THE GENETIC BASIS FOR PIGMENT VARIATION
AMONG GREEN SULFUR BACTERIA

A Thesis in
Biochemistry and Molecular Biology
by
Julia A. Maresca

Submitted in Partial Fulfillment
of the Requirements
for the Degree of
Doctor of Philosophy

May 2007
The thesis of Julia A. Maresca was reviewed and approved by the following committee members*

Donald A. Bryant  
Ernest C. Pollard Professor of Biotechnology and Professor of Biochemistry and Molecular Biology  
Thesis Advisor  
Chair of Committee

John H. Golbeck  
Professor of Biochemistry and Biophysics

Sarah E. Ades  
Assistant Professor of Biochemistry and Molecular Biology

Squire J. Booker  
Associate Professor of Biochemistry and Molecular Biology

Lee R. Kump  
Professor of Geosciences

Robert A. Schlegel  
Professor of Biochemistry and Molecular Biology  
Department Head, Department of Biochemistry and Molecular Biology

*Signatures are on file with the Graduate School
Abstract

The pigmentation differences between green-colored and brown-colored green sulfur bacteria (GSB) are more than cosmetic: species with different pigmentation inhabit different parts of the photic zone. Green-colored species, which make bacteriochlorophyll (BChl) c or d as their primary antenna BChl and chlorobactene as their main carotenoid, tend to be found in the upper layer of anaerobic photic zones. Brown-colored species, which make BChl e and the dicyclic carotenoid isorenieratene, are usually found deeper in the water column. These pigment pairs are invariant, which means that for a green species to become a brown one or vice versa, changes in two unrelated biosynthetic pathways must occur. In this work, comparative genomics has been used to identify the genes unique to pigment biosynthesis in green-colored and brown-colored green sulfur bacterial species. The gene encoding the C-20 methyltransferase, responsible for the difference between BChls c and d, has been identified and inactivated, and detailed analysis of BChl c- and d-producing strains has provided a molecular explanation for the observation that BChl c-containing species tend to live in darker environments. Additionally, a cluster of genes found exclusively in the genomes of brown-colored GSB species has been identified, and fragments of this cluster have been inserted into the chromosome of Chlorobium (Chl.) tepidum to investigate their role in BChl e biosynthesis. Using phylogenetic profiling, carotenoid cyclases specific to chlorobactene or isorenieratene biosynthesis were identified. These cyclases are responsible for the difference between mono- and dicyclic carotenoids, and their activity has been characterized both in GSB and in a heterologous expression system in Escherichia coli. Based on these analyses, the biosynthetic pathway for chlorobactene was established and a similar pathway for isorenieratene biosynthesis is demonstrated. Lastly, analysis of genome regions has identified chlorobactene-modifying enzymes which synthesize the membrane-associated OH-chlorobactene acyl glycosides in both green-colored and brown-colored GSB. These genes have been inactivated in Chl. tepidum and their roles in synthesizing glycosylated and acylated carotenoids have been confirmed. The investigations described in this work explain the genetic basis for stratification of green-colored species in the environment, complete the biosynthetic pathway for chlorobactene, identify the first known isorenieratene-specific carotenoid cyclase, and explain some of
the natural variation in carotenoid end products seen in different species of green sulfur bacteria. Four of these proteins, the carotenoid cyclases, the carotenoid glycosyltransferase, and the carotenoid acyltransferase, are the first-characterized members of what appear to be large families of carotenoid-modifying enzymes.
Table of Contents

List of Tables x
List of Figures xi
Acknowledgements xiv

CHAPTER 1. Introduction. 1

1.1 Green sulfur bacteria 2
1.2 Phylogeny of green sulfur bacteria 2
1.3 Ecology of green sulfur bacteria 4
1.4 Carbon and sulfur metabolism 5
1.5 Bacteriochlorophylls in the chlorosome: the antenna of green sulfur bacteria 6
1.6 (Bacterio)chlorophylls associated with the chlorosome baseplate, the FMO protein, and the photosynthetic reaction center. 8
1.7 Carotenoids in green sulfur bacteria 9
1.8 Genome sequences of green sulfur bacteria 10
1.9 Identification of genes unique to pigment biosynthesis in green sulfur bacteria 12

CHAPTER 2. Identification of a new class of carotenoid cyclases in photosynthetic organisms 38

Abstract 39

2.1 Introduction 40

2.2 Methods

2.2.1 Bacterial strains and growth conditions 43
2.2.2. Phylogenetic profiling 43
2.2.3 Preparation of genomic library and identification
2.2.4 Construct for inactivation of cruA (CT0456) in Chl. tepidum

2.2.5 Pigment analysis

2.3 Results

2.3.1 Phylogenetic profiling

2.3.2 Identification of lycopene cyclase in Chl. tepidum

2.3.3 Inactivation of Lycopene Cyclase in Chl. tepidum

2.3.4 Phylogenetic Analyses of Lycopene Cyclases

2.4 Discussion

2.5 References

CHAPTER 3. Heterologous expression of CruA-type carotenoid cyclases and proposed biosynthetic pathway for isorenieratene

Abstract

3.1 Introduction

3.2 Methods

3.2.1 Strains and growth conditions

3.2.2 Constructs for expression of CruA and CruB in E. coli

3.2.3 Assays of CruA and CruB activity in E. coli

3.2.4 Expression of cruB in Chl. tepidum

3.2.5 Inhibition of carotenoid cyclization

3.2.6 Pigment analysis

3.3 Results
Abstract 129

5.1 Introduction 130

5.2 Materials and Methods

5.2.1 Strains and growth conditions 132

5.2.2 Inactivation of CT0028 132

5.2.3 Chlorosome preparation and analysis 132

5.2.4 Growth rate measurements 133

5.2.5 Competition experiments 133

5.3 Results

5.3.1 Identification of genes potentially encoding the BChl c C-20 methyltransferase 135

5.3.2 Sequence analysis of bchU and crtF genes 136

5.3.3 Construction and verification of a CT0028 mutant of Chl. tepidum 137

5.3.4 Pigment analysis of Chl. tepidum CT0028 (bchU) mutant 137

5.3.5 Analysis of molar extinction coefficients of BChl c and d 138

5.3.6 Growth characteristics of BChl c and BChl d strains 139

5.3.7 Competition between BChl c- and d-containing strains 139

5.4 Discussion 141

5.5 References 146

CHAPTER 6: Oxidation of the C71 position of bacteriochlorophyll e 166

Abstract 167
List of Tables

Chapter 2.

Table 2.1 Primers used
Table 2.2 Protocols for HPLC analysis of pigments

Chapter 3.

Table 3.1 Primers used in this work
Table 3.2 Plasmid combinations in *E. coli* strain BL21(DE3): pAC-LYC
Table 3.3 Plasmid combinations in *E. coli* strain BL21(DE3): pAC-NEUR

Chapter 4.

Table 4.1 Primers used in this work

Chapter 5.

Table 5.1 Absorption maxima *in vivo* and in methanol of bacteriochlorophylls c, d, and e.
Table 5.2 Primers used in this work
Table 5.3 Growth rates of bacteriochlorophyll c- and d-producing strains of *Chl. tepidum* and *Chl. vibrioforme*

Chapter 6.

Table 6.1 Primers used in this work
Table 6.2 Genes possibly specific to BChl e biosynthesis in the region around *cruB* in the genomes of 3 brown-colored GSB
Table 6.3 Distribution of genes *BE1* and *cruB* among green-colored and brown-colored GSB species in a culture collection
List of Figures

Chapter 1.

Figure 1.1  Phylogeny of the major eubacterial taxa based on RecA sequences
Figure 1.2  Phylogeny of GSB based on 16S rDNA sequences
Figure 1.3  Micrograph of Chl. tepidum cells and extracellularly deposited S^0
Figure 1.4  Transmission electron micrograph of Chl. tepidum cell, showing chlorosomes.
Figure 1.5  Diagram of the chlorosome and transfer of excitation energy to the reaction center
Figure 1.6  Structures of the antenna bacteriochlorophylls found in GSB
Figure 1.7  Diagram of the GSB photosynthetic reaction center, showing the path of electron transfer
Figure 1.8  Structure of bacteriochlorophyll a
Figure 1.9  Structure of chlorophyll a
Figure 1.10 Structures of carotenoids found in green-colored species of GSB
Figure 1.11 Structures of some carotenoids found in brown-colored species of GSB

Chapter 2.

Figure 2.1  Two bacterial pathways for the synthesis of lycopene
Figure 2.2  Isomerization of the linear ψ end group to cyclic end groups
Figure 2.3  Phylogenetic profiles of photosynthesis-related ORFs
Figure 2.4  Elution profiles and absorption spectra of carotenoids produced in E. coli screen
Figure 2.5  Plasmid pCTLY
Figure 2.6  Plasmid pCTLY-aadA and segregation of CT0456 mutant
Figure 2.7  Elution profiles of carotenoids from wild-type Chl. tepidum and CT0456::aadA mutant
Figure 2.8  Phylogenetic tree of all known lycopene cyclases
Figure 2.9  Biosynthetic pathway of chlorobactene in Chl. tepidum
Figure 2.10 Isomerization of isopentenyl diphosphate to dimethylallyl diphosphate by IPP isomerase

Chapter 3.

Figure 3.1 Structures of chlorobactene and isorenieratene
Figure 3.2 Phylogenetic tree of cruA and cruB from GSB
Figure 3.3 Plasmids pCPL1 and pCOLA-B
Figure 3.4 Plasmid pXN-BEBj-aadA
Figure 3.5 Elution profiles and absorption spectra of carotenoids produced in E. coli BL21(DE3) pAC-LYC expressing cruA and/or cruB
Figure 3.6 Elution profiles and absorption spectra of carotenoids produced in E. coli BL21(DE3) pAC-NEUR expressing cruA and/or cruB
Figure 3.7 Elution profile of carotenoids from Chl. tepidum expressing cruB.
Figure 3.8 Elution profile of pigments from Chl. tepidum, Chl. phaeobacteroides strain 1549, and Pld. phaeum supplemented with MPTA
Figure 3.9 Pigments produced in E. coli BL21(DE3) pAC-LYC + pCPL1 treated with MPTA
Figure 3.10 Proposed biosynthetic pathway of isorenieratene in Pld. phaeum and Chl. phaeobacteroides strain 266T
Figure 3.11 Proposed biosynthetic pathway of carotenoids in Chl. phaeobacteroides strain 1549

Chapter 4

Figure 4.1 Biosynthetic pathway of hydroxychlorobactene-glucoside laurate
Figure 4.2 Genome regions of CT1987 and CT0967, and regions surrounding their homologs in other carotenogenic species.
Figure 4.3 Constructs for inactivation of CT1987 and CT0967 and segregation of the mutants.
Figure 4.4  HPLC analysis of carotenoids produced by wild-type *Chl. tepidum* and mutants unable to synthesize OH-chlorobactene glucoside laurate

Figure 4.5  Growth rates of wild-type *Chl. tepidum* and mutants unable to synthesize OH-chlorobactene glucoside laurate

Figure 4.6  Phylogenetic tree of carotenoid glycosyltransferases

Figure 4.7  Phylogenetic tree of acyltransferases

**Chapter 5**

Figure 5.1  Structures of bacteriochlorophylls *c*, *d*, and *e*

Figure 5.2  Photosynthesis-related gene cluster in *Cfx. aurantiacus* and phylogenetic tree of BchU and CrtF sequences

Figure 5.3  Alignment of *bchU* sequences

Figure 5.4  *bchU* inactivation construct and confirmation of segregation of *bchU* mutant

Figure 5.5  Elution profiles of bacteriochlorophylls from wild-type *Chl. tepidum* and from *bchU* mutant

Figure 5.6  Molar absorption coefficients of bacteriochlorophylls *c* and *d* in methanol and in aqueous buffer

Figure 5.7  Growth rates of *Chl. tepidum* strains

Figure 5.8  Growth rates of *Chl. vibrioforme* strains

Figure 5.9  Competition experiments with *Chl. tepidum* strains

Figure 5.10  Competition experiment with *Chl. vibrioforme* strains at 70 µmol photons m\(^{-2}\) s\(^{-1}\)

**Chapter 6**

Figure 6.1  Polyphyletic distribution of pigment phenotypes among GSB

Figure 6.2  Structures of bacteriochlorophylls *c* and *e*

Figure 6.3  Plasmids used for creating *Chl. tepidum* strains capable of expressing putative BChl *e*-specific genes from chromosomal loci
Figure 6.4  Results of reciprocal BLAST hit analysis of GSB genomes

Figure 6.5  Gene neighborhoods around cruB in three brown-colored species of GSB

Figure 6.6  Phylogenetic relationships between green sulfur bacterial bchF homologs

Figure 6.7  Phylogenetic tree of selected proteins in the radical SAM family

Figure 6.8  Elution profiles of BChls from wild-type Chl. tepidum and three isolates of the csmH::nBE-aadA strain

Figure 6.9  Elution profiles of BChls from wild-type Chl. tepidum, wild-type Pld. phaeum, and strain csmH::PXE-aadA

Figure 6.10 Elution profiles of BChls from wild-type Chl. tepidum and three isolates of the bchU::nBEDH-aadA strain
Acknowledgements

Many thanks first to Donald A. Bryant, for his encouragement and advice and for the many opportunities he gave me that I would not have had elsewhere. Thanks also to Niels-Ulrik Frigaard, who began both of the pigment biosynthesis projects in Chlorobium tepidum and trained me in the techniques necessary to continue these projects. Joel E. Graham and Aline Gomez Maqueo Chew were invaluable colleagues, and their assistance and ideas improved my work significantly. Without Tao Li, who closed the green sulfur bacterial genome sequences, the genome comparisons would have been incomplete. Christopher H. House (Department of Geosciences) helped me with the coding for the whole-genome comparisons, and Jorg Overmann (Ludwig-Maximilians University, Munich, Germany) made his culture collection available for analysis.

Funding for the pigment biosynthesis work was provided by the U.S. Department of Energy, grant # DE-FG02-94ER20137 to Donald A. Bryant. Funding for the work done in Jorg Overmann’s lab as well as for the Sulfolobus solfataricus project was provided by the Biogeochemistry Research Initiative in Education, an NSF-IGERT grant awarded to the Department of Geosciences at the Pennsylvania State University.
CHAPTER 1

Introduction to green sulfur bacteria.
1.1 Green sulfur bacteria

Green sulfur bacteria (GSB) are obligately anaerobic, photolithoautotrophic bacteria found in sulfidic, anoxic photic zones. They are part of a very small subset of organisms that can live solely on inorganic substrates: all known GSB species fix CO$_2$ and N$_2$, using electrons derived from reduced sulfur or reduced iron compounds. As the organisms responsible for most carbon fixation in anaerobic environments, they are of interest to those studying early microbial life, those studying life in extreme environments, and those studying global carbon cycling.

1.2 Phylogeny of green sulfur bacteria

Because both GSB and the filamentous anoxygenic phototrophic bacteria (FAPs) both make chlorosomes as their photosynthetic antenna (24, 88), it was assumed that they were related to each other. When molecular techniques for phylogenetic analysis were developed, no evidence for a phylogenetic relationship between these two groups could be shown (43). Based on molecular phylogenies using 16S rDNA (107, 108) and RecA protein sequences (52; see Figure 1.1) as well as phylogenies based on whole-genome analysis (23; T. Li, personal communication), the GSB clade is the sister group of the Cytophaga-Flexibacteria-Bacteroides (CFB) clade, and is only distantly related to all other phototrophic bacterial species.

The relationship of the green sulfur bacterial photosynthetic apparatus to that of other phototrophs is more difficult to describe. Like cyanobacteria, GSB have Type I (iron-sulfur-type) reaction centers (RCs), which generate a strong reductant, reduced ferredoxin. The genes used for production of carotenoids in GSB are also very similar to those of cyanobacteria (38, 92; see Chapter 2). The biosynthetic pathway of bacteriochlorophyll (BChl) $c$ biosynthesis is similar to the pathways for Chl $a$ and BChl $a$ biosynthesis in cyanobacteria and proteobacteria (48). However, the photosynthetic antenna of GSB, the chlorosome, is found only in GSB, FAPs, and the recently-identified phototrophic acidobacterium Candidatus “Chloracidobacterium thermophilum” (16). The Fenna-Matthews-Olson (FMO) protein, which lies between the chlorosome and the photosynthetic reaction centers in GSB, has only been found in GSB and Candidatus “Chloraci-
dobacterium thermophilum” (16). Thus, it is impossible to conclude how GSB are related to other phototrophs based on the components of the photosynthetic apparatus.

Phylogenies of photosynthetic bacteria, reconstructed for the purpose of understanding the evolution of photosynthesis, have yielded contradictory results. Analysis of amino acid sequences for proteins unrelated to photosynthesis from photosynthetic bacteria results in phylogenies similar to those based on 16S rDNA and RecA (e.g. 51, 53). A phylogeny of Type I RCs, based on a 159-amino-acid fragment of “core antenna domains” (PscA, PshA, PsaA, PsaB, PsbB, PsbC, IsiA, and Pcb), has concluded that these proteins were present in the last common ancestor of cyanobacteria, heliobacteria, and GSB, and were inherited vertically (96). In contrast, the work of Xiong et al. (151), in which the amino acid sequences of proteins required for (B)Chl biosynthesis, carotenoid biosynthesis, charge separation, and biosynthesis of other cofactors were concatenated in different groups, concludes that the closest relatives of GSB are the FAPs and that the Gram-positive heliobacteria are more related to cyanobacteria than are GSB. The presence of chlorosomes in groups as diverse as GSB, FAPs, and acidobacteria (Candidatus “Chloracidobacterium thermophilum,” [16]) suggests that the biosynthetic pathways for BChl c and chlorosome biogenesis may have been horizontally transferred more than once (11, 120). Additionally, it has been noted that carotenoid biosynthesis genes – and occasionally entire pathways – have been laterally transferred many times (39, 45, 91, 95, 112; see Chapter 4). Logically, it is almost impossible to infer the evolutionary history of photosynthesis, as a process, based on individual protein sequences or even on individual pathways, because lateral gene transfer of different pathways or pathway fragments, as well as loss of the genes, must have happened several times, in several different lineages (11, 102, 150).

Within the GSB clade, four genera exist (Figure 1.2). The nomenclature describing GSB was recently amended to reflect molecular phylogenies rather than the previously used metabolic and morphological descriptions (67); in this work, the old names will be used. The first group has only one known cultivated member, Chloropheretomon thalassium (42), and is the most divergent genus (1). Genes encoding 16S rRNA, the FMO protein, and PscA from a putative GSB have been sequenced from Octopus Spring in Yellowstone National Park (D.A. Bryant, personal communication), and based on
these sequences, this putative GSB seems to belong to the *Chloroherpeton* group. Because of the limited genetic data available and the absence of a cultured organism, its phylogenetic affiliation is not settled. The other three genera are *Chlorobium*, *Chlorobaculum*, and *Prosthecochloris*. These groups are easily distinguished using molecular phylogenetic methods (1, 67, 107). However, aside from the high salt requirements of *Prosthecochloris* spp., no obvious morphological or metabolic characteristics distinguish these genera from each other, and in fact some *Chlorobium* and *Chlorobaculum* species also require high-salt growth conditions (67).

The species *Chlorobium tepidum* TLS, referred to as *Chlorobaculum tepidum* in the new taxonomic system, is the type strain for the new genus *Chlorobaculum* (67). This species has become the model organism for GSB because it grows rapidly, tolerates high light intensity, and is genetically amenable (33, 146). This species was used as the basis for all of the genetic work described in this thesis. The other species used for physiological experiments in this work are the closely-related *Chl. vibrioforme* (now *Chlorobaculum parvum*) strain NCIB 8327, a BChl *d*-producing, green-colored species; *Chl. phaeobacteroides* (now *Chlorobaculum limnaeum*) strain 1549, a brown-colored, BChl *e*-producing species with an unusual complement of carotenoids (61); and *Pelodictyon phaeum*, which has not yet been renamed, a brown-colored, BChl *e*-producing species which synthesizes isorenieratene. All of these species can, like *Chl. tepidum*, utilize thiosulfate as an electron donor and grow at temperatures of 28-30°C (*Chl. tepidum* grows optimally at 48°C).

### 1.3 Ecology of green sulfur bacteria

The requirement for sulfide (or reduced iron) and light and their sensitivity to oxygen restricts GSB to a narrow range of environments. They are found below the chemocline in stratified lakes (e.g. 133, 137, 143), in shallow anoxic sediments (42), and in sulfidic springs (148). GSB have also been isolated from below the chemocline of the Black Sea (90, 105) and from the water close to a hydrothermal vent more than 2000 m below the surface of the Pacific Ocean (9). The *in situ* temperatures in these environments range from 2°C in the ocean around the hydrothermal vent (9) to 68°C in the Octopus Spring microbial mats in Yellowstone National Park (30), though the highest temperature at
which a cultivated GSB grows optimally is 48°C for Chl. tepidum (148). GSB tend to
grow best at neutral pH (6.8-7.0; 104), but have been detected in environments with pH
values as low as 2.5 (26) and Chl. tepidum was isolated from a spring with a pH of 4
(148).

GSB have been found in marine, aquatic, and benthic environments. In aquatic
systems, the green-colored species of GSB are usually found in shallower locations than
the brown-colored species, where the red wavelengths of light are relatively enriched (65,
99, 105, 141, 142). The brown-colored species, on the other hand, are frequently found
either in very deep water (90, 105) or in lakes that have dense populations of PSB at the
chemocline (41, 141, 142). Both of these environments are enriched in blue to blue-green
wavelengths (27, 137). Most of the GSB isolated from sediments have been green-
...
carbon derived from a reverse TCA pathway of carbon fixation, in conjunction with identification of the GSB-specific biomarker isorenieratane and its derivatives, has enabled researchers to conclude that certain environments, at least temporarily, had anoxic photic zones (e.g. 69, 136).

All but one known species of GSB require a source of reduced sulfur as an electron donor (104); of these, all utilize sulfide and several can use thiosulfate or elemental sulfur (S\textsuperscript{0}) as electron donors. Sulfide is oxidized to zero-valent sulfur in a 2-electron transfer mediated by flavocytochrome c (13, 14) or sulfide:quinone reductase (SQR; 126); the resulting S\textsuperscript{0} granules are deposited extracellularly (Figure 1.3). When the sulfide is depleted, the polysulfur granules may be oxidized further to sulfite by dissimilatory sulfite reductase (DSR). Although GSB oxidize sulfide to sulfate, the enzymes catalyzing oxidation of elemental sulfur to sulfite have not been characterized biochemically. This pathway of sulfur oxidation is inferred from the fact that all of the available genome sequences encode homologs to the *dsr* genes from the purple sulfur bacterium (PSB) *Allochromatium vinosum*, which are required by that organism for oxidation of intracellular elemental sulfur (35, 123). Oxidation of sulfite to sulfate is carried out by APS reductases in at least two thiosulfate-utilizing species (72, 74) and probably in *Chl. tepidum*, *Chl. chlorochromatii*, *Pelodictyon (Pld.) phaeoclathratiforme*, and *Chl. phaeobacteroides* strain BS1, whose genomes all encode homologs of the APS reductase (35).

1.5 Bacteriochlorophylls in the chlorosome: the antenna of green sulfur bacteria

Because they are obligate phototrophs living in light-limited environments, GSB require an extremely efficient light-harvesting antenna. Their antenna, the chlorosome, is the largest antenna known (36). Chlorosomes are membrane-bound compartments underlying the cytoplasmic membrane (24; Figure 1.4) with arrays of self-aggregated bacteriochlorophyll (BChl) *c*, *d*, or *e* (63). A complex consisting of the Fenna-Matthews-Olson (FMO) protein and BChl *a* (101) lies between the chlorosome and the photosynthetic reaction centers, which are in the cytoplasmic membrane. Chlorosomes are also found in FAPs (88) and the acidobacterium *Candidatus “Chloracidobacterium (Cab.) thermophiliun”* (16). Although these bacteria have thus far been found in microbial mats rather than
deep aquatic systems (16, 31, 70, 77, 113), the light gradients in these mats are so steep that the light intensity can be less than 1% of the surface irradiance within a few mm of the surface (31), and a large antenna such as a chlorosome would provide an advantage to the phototrophs there. The chlorosomes in FAPs are smaller than those in GSB and lack the FMO-BChl a complex between the chlorosomes and the reaction centers (40), while the chlorosomes in Cab. thermophilum are similar in size to those of GSB and have an FMO-BChl a complex (16). The existence of this structure in such distantly related organisms suggests that its biogenesis may require only a few genes other than those required for BChl synthesis.

The chlorosome envelope in *Chl. tepidum* has 10 proteins associated with it (22, 37; Figure 1.5). One of these, CsmA, is a 6.2-kDa protein that binds a single BChl a and forms multimers in the baseplate of the chlorosome (22, 17, 122). The *csmA* gene cannot be inactivated, indicating that its function is essential for viability (37). The genes encoding the other nine chlorosome proteins can be inactivated in *Chl. tepidum* with no effect on growth or photosynthetic efficiency (37), and analysis of the chlorosome shapes of different double mutants has indicated that these proteins play a role in establishing the size, shape, and redox properties of the chlorosome (82).

The antenna chlorophylls in the chlorosomes of GSB may be one of the three “*Chlorobium* chlorophylls,” BChl c, d, or e (Figure 1.6), which differ at the C7 and C20 positions. BChl d has a methyl group at the C7 position (118), BChl c has a methyl group at the C7 position and an additional methyl at the C20 position (62), and BChl e is methylated at the C20 position and formylated at the C7 position (46). The hypothetical pigment BChl f, which would be formylated at the C7 position and unmethylated at the C20, has been synthesized and analyzed *in vitro* (131) but has never been observed in nature. The antenna BChl dictates what kind of environmental niche a species may inhabit (see section 1.3); the BChl e-producing species tend to be found in darker environments enriched in blue wavelengths of light, and the BChl c- or d-producing species tend to live in environments enriched in red and near-infrared wavelengths.

The antenna BChls aggregate into large three-dimensional arrays without a protein scaffold (8, 138). The BChls in GSB have a hydroxyl group at the C3\(^1\) position, which is unique among (B)Chls. The C3\(^1\) hydroxyl on one BChl interacts with the mag-
nesium on an adjacent BChl, whose C3\(^1\) hydroxyl interacts with another BChl; in this way the BChls form large-scale aggregates in non-polar environments (59, 60). In fact, each chlorosome may contain up to 215,000 molecules of BChl c, d, or e (97). Since each GSB cell has 200 to 250 chlorosomes, each cell can have as many as 53.7 \times 10^6 BChl molecules. Unlike all other photosynthetic antenna, the BChls in the chlorosome do not require a protein scaffold to orient them in the proper position; thus, this antenna requires less energy to synthesize than other antennas.

1.6 (Bacterio)chlorophylls associated with the chlorosome baseplate, the FMO protein, and the photosynthetic reaction center.

The energy from a photon is absorbed by the aggregated antenna BChl and quickly transferred to the CsmA-BChl \(a\) complex in the baseplate of the chlorosome. CsmA has been shown to form a paracrystalline array (128), and cross-linking studies have demonstrated the existence of multimers with more than 8 subunits (83). The absorption maximum of the BChl \(a\)-CsmA complex, 798 nm, is red-shifted relative to the absorption maxima of the antenna BChl aggregates (98, 122); thus, the excitation energy is efficiently transferred from the antenna to the baseplate (21).

The absorption maximum of the BChl \(a\)-FMO protein complex between the CsmA-BChl \(a\) complexes and the photosynthetic reaction centers is red-shifted relative to the CsmA complex, with an absorption maximum at 808 nm (57). The FMO protein has been crystallized (18, 85, 94, 134) as a homotrimer, with each monomer binding 7 molecules of BChl \(a\). The FMO protein shares a short motif with the GSB reaction center protein PscA, and this has been interpreted to mean that these proteins are distantly related to each other (103).

Both BChl \(a\) and Chl \(a\) are found in the reaction center, in which BChl \(a\) is the P840 primary donor, and Chl \(a\) is the primary acceptor \(A_0\) (29; Figure 1.7). In contrast to BChls c, d, and e, BChl \(a\) is reduced at the C7-8 bond, has a keto group rather than a hydroxyl at the C3\(^1\) position, and has a C13\(^2\) carboxymethyl group (Fig. 1.10; 12). These structural changes shift the \(Q_y\) absorption band of BChl \(a\) to 770 nm in solvents, and the protein-pigment interactions shift it even further. Chl \(a\), unlike either BChl \(a\) or the an-
tenna BChls, has a vinyl group at the C3 position; like BChl a, it has a C13\textsuperscript{2} carboxymethyl group, and like BChl c has an oxidized C7-8 bond (Figure 1.9).

When the excitation energy from a photon reaches the primary donor, P840, charge separation occurs and an electron is transferred from P840 to A\textsubscript{0}, which is Chl a (29). From Chl a, the electron is transferred to the interpolypeptide [4Fe-4S] center F\textsubscript{X} (140); from F\textsubscript{X} the electron is transferred to the F\textsubscript{A} and F\textsubscript{B} Fe-S clusters in the PscB protein (76, 139), and from there to ferredoxin (57). Menaquinone-7 has been reported to be the electron acceptor A\textsubscript{1} in GSB reaction centers (75); however, in other reports, evidence for the involvement of this quinone in electron transfer has not been found (80, 139, 140).

1.7 Carotenoids in green sulfur bacteria

Most of the carotenoids in GSB are located in the chlorosome (117), where they affect the aggregation of the antenna BChl (78, 115, 116), participate in light-harvesting and photoprotection (20), and play a role in protection from exposure to oxygen (84). The carotenoids produced by GSB are primarily the monocyclic aryl compound chlorobactene (87; Figure 1.10) in green-colored species, and the dicyclic, di-aryl compound isorenieratene (86) in brown-colored species (Figure 1.11). Additionally, both green-colored and brown-colored species of GSB produce small amounts of acylated OH-chlorobactene glucosides (38, 61, 86, 130).

Chl. tepidum is a green-colored GSB species, which makes chlorobactene as its primary carotenoid (38, 137). It also makes significant amounts of γ-carotene, 1',2',dihydrochlorobactene, 1',2',dihydro-γ-carotene, and glucoside esters of hydroxychlorobactene and hydroxy-γ-carotene (38, 137; Figure 1.10). The precise functions of each of these end products is still unclear; inhibiting carotenoid biosynthesis at any step after isomerization of poly-cis-lycopene has very little effect on growth rates (38; see chapter 2), although preventing synthesis of the glycoside esters does reduce the growth rate somewhat (38, 91; see chapter 4).

If carotenoid biosynthesis in Chl. tepidum is stopped by inactivation of any of the first four genes in carotenoid biosynthesis, growth is severely impaired under all conditions tested (38). These mutants also have decreased BChl a content, and in light of the
fact that the CsmA-BChl a chlorosome baseplate complex in *Cfx. aurantiacus* has approximately 2.5 carotenoid molecules per CsmA monomer (98), these results suggest that carotenoids may be important in protecting BChl a in the baseplate. Recent analysis of chlorosomes from a *Chl. tepidum* mutant lacking all colored carotenoids has shown that the structures of the BChl c aggregates in the chlorosomes of this mutant are significantly disrupted (66). Together, these results suggest that both photoprotective mechanisms and light-harvesting efficiency are impaired when colored carotenoids are absent (38; see Appendix B).

No genetic system exists for inactivation of genes in a brown-colored GSB species (see Appendix C). However, synthesis of colored carotenoids in these species can be inhibited with 2-hydroxybiphenyl (HBP), a specific inhibitor of phytoene desaturation (32). When the brown-colored GSB *Chl. phaeobacteroides* strain CL 1401 was treated with HBP, the cells accumulated the colorless phytoene (Figure 1.11) instead of isorenieratene (5, 6). The Q<sub>y</sub> absorption maximum *in vivo* was slightly red-shifted, and both CsmA and BChl a were reduced by ~40% (6). In addition, the chlorosomes in HBP-treated cells were smaller and more rounded than chlorosomes from untreated cells (5). Again, these results suggest that colored carotenoids play a role in the aggregation of the antenna BChl as well as stabilizing the CsmA-BChl a baseplate, which dictates the footprint size of the chlorosome (6).

### 1.8 Genome sequences of green sulfur bacteria

After manual annotation, the sequenced genome of the moderately thermophilic GSB species *Chl. tepidum* was made available in 2002 (28). Based on genes found in this genome and in the draft genome sequence of the unrelated FAP *Chloroflexus* (*Cfx.*) *aurantiacus*, which also makes BChl c and chlorosomes, predictions about BChl biosynthesis, carotenoid biosynthesis, and chlorosome biogenesis were made (15, 28, 34, 36) and the genes predicted to be important in these three pathways were systematically inactivated and analyzed (36, 37, 38, 47, 48, 49, 82, 91, 92, 93; see Chapters 2, 4, and 5).

A single genome, however, can only provide limited information, and *Chl. tepidum* is an atypical GSB in its optimal growth temperature, its ability to utilize thiosulfate as an electron donor, its ability to grow at high light intensities, and its rapid growth
rate. To extend the available sequence information, eleven more GSB genomes have been sequenced, together encompassing species with different electron donor requirements, morphologies, environmental niches, and pigmentation and including at least one member of each of the GSB clades. These include *Chl. chlorochromatii*, the green sulfur bacterial epibiont of the consortium “Chlorochromatium aggregatum” (144); *Chl. ferrooxidans*, which uses reduced iron rather than reduced sulfur as an electron donor (58); *Chl. limicola* strain DSM 245\(^T\), the first described species of GSB and one which, like *Chl. tepidum*, was isolated from a hot spring (100); the brown-colored, BChl \(\epsilon\)-producing aquatic species *Chl.phaeobacteroides* strain DSM 266\(^T\) (111); *Chl. phaeobacteroides* strain BS1, a brown-colored strain recovered from below the chemocline of the Black Sea, where the light intensity is less than 3 nmol photons m\(^{-2}\) s\(^{-1}\) (90, 105); *Chl. vibrioforme* strain 265\(^T\) (now called *Prosthecochloris vibrioformis*), also a high-salt-tolerant species (110); the BChl \(d\)-producing strain *Chl. vibrioforme* strain NCIB 8327d (*Chlorobaculum parvum*), which is also genetically amenable and can utilize thiosulfate as an electron donor; *Pelodictyon (Pld.) luteolum* strain DSM 273\(^T\), a green-colored, gas vacuolated species which may form clumps or hollow colonies, first identified as algae in 1901 (125); *Pld. phaeoclathratiforme* strain BU-1, a brown-colored species which makes gas vesicles and forms net-like filaments (106); *Prosthecochloris aestuarii* strain DSM 271\(^T\), isolated from marine sediments and forming prosthcae (50); and the species *Chloroherpeton thalassium*, isolated from marine sediments (42).

These genomes represent a very large data set about a small group of closely related of organisms. Because they are so closely related and vary in defined ways, such as salt tolerance, thiosulfate utilization, production of gas vesicles, formation of colonies or filaments, or pigment content, comparisons of these genomes should enable identification of genes involved in each of these functions, as well as a conserved core of GSB-specific genes. In fact, preliminary comparisons of these genomes with each other and with other genomes has already enabled predictions about the proteins involved in sulfur or iron metabolism in GSB, as well as how these genes might be transferred between species (35).
1.9 Identification of genes unique to pigment biosynthesis in green sulfur bacteria

The pigments made in GSB include the “Chlorobium” chlorophylls BChl c, d, and e, the carotenoids γ-carotene, chlorobactene, β-carotene, β-isorenieratene, and isorenieratene, and the carotenoid derivative OH-chlorobactene glucoside laurate. The purpose of the work described here was to identify the gene(s) unique to the biosynthesis of each of these pigments. In the carotenoid biosynthetic pathways, these genes are carotenoid cyclases, one specific to chlorobactene biosynthesis and one specific to isorenieratene synthesis. The synthesis of glycosylated and acylated chlorobactene is a branch of carotenoid biosynthesis common to all GSB, and the hydroxychlorobactene glycosyltransferase and acyltransferase remain to be identified. In the BChl pathways, the genes responsible for variation include the BChl c- and e- specific C20 methyltransferase and the gene(s) responsible for oxidation of the C7¹ carbon to an aldehyde in BChl e-producing species.

Identification of the genes responsible for the biosynthesis of these signature pigments will have several results. Better understanding of the biosynthetic pathways for chlorobactene, isorenieratene, and BChls c, d, and e in GSB may clarify controversies regarding the evolution of Chl biosynthesis (48) and distribution of different carotenoid biosynthetic genes among bacteria. Identification of the isorenieratene- and BChl e-specific genes may give us enough information to propose a genetic explanation for the polyphyletic nature of the brown-colored and green-colored phenotypes, both of which appear in each major GSB clade. Lastly, analysis of the mutant strains producing different pigments as end products should enable us to provide a molecular explanation for the vertical distribution trends of GSB in the environment (see section 1.1 above).

Identifying a gene with a specific function is less difficult when one or more genome sequences are available. If only one genome sequence is available, the predicted ORFs in that genome can be analyzed and ORFs homologous to proteins with similar functions, ORFs with domains shown to catalyze similar reactions, or ORFs with recognizable substrate- or cofactor-binding motifs can be identified as candidate genes. This kind of analysis, however, is not useful in the case of ORFs that are merely conserved hypothetical proteins (CHPs) with no identifiable homologs or motifs, and it may generate a long list of false positive results (49, 93; see chapter 5). Analysis of the neighborhoods of the genes of interest may make the list of candidate genes more specific, as it...
has in other cases (7, 89, 145, 149). In bacterial genomes, genes whose products are involved in the same pathway or located in the same compartment are frequently grouped together. Thus, if a gene encoding a CHP is close to several genes known to be important for carotenoid biosynthesis, it becomes more likely that that gene is part of the same pathway. This analysis is only useful in the case of organisms known to co-regulate transcription of genes in the pathway of interest; GSB appear to regulate very few pathways at the level of transcription (28), so gene neighborhood analysis would be more useful in comparing GSB genomes to genomes from other bacterial lineages.

Another useful tool for whole-genome analysis is phylogenetic profiling. In this type of analysis, presence or absence of a gene in a genome is matched with presence or absence of a trait in an organism (109). Genes found in organisms that have that trait – and, equally importantly, not found in organisms that do not have it – are more likely to be involved in similar processes. This is an especially useful tool for comparing genomes of organisms that share very specific traits, such as GSB and cyanobacteria, which have in common a carotenoid biosynthesis pathway. Similar analyses can be used to compare the genomes of green-colored and brown-colored GSB; because the GSB genomes are, on average, very small and highly conserved between species, the genes unique to and shared by the brown-colored species have a high probability of being involved in pigmentation.

Using the available genome sequences to identify candidate genes for specific steps in pigment biosynthesis, and then using the available genetic system in Chl. tepidum to confirm their functions, will clarify which types of genome analysis are most effective at predicting genes with particular functions in GSB. Many papers have appeared in recent years that predict the functions of uncharacterized genes (54, 109, 121), but relatively few of them have used this information as the basis for experiments (25, 81, 132). In addition, characterization of this set of genes will complete several biosynthetic pathways. Because the products of these pathways enable GSB to harvest light in environments with very low light fluxes, identification of the genes involved in synthesis of their pigments will contribute to our understanding of how GSB have adapted to perform photoautotrophy in such niches.
Figure Legends

**Figure 1.1** Phylogenetic tree of selected bacterial species based on RecA amino acid sequences. Asterisks indicate lineages with photosynthetic members.

**Figure 1.2** Phylogenetic tree of green sulfur bacteria based on 16S rDNA sequences. Branches with dark diamonds indicate brown-colored species and light diamonds indicate green-colored species. After Alexander *et al.*, 2002 (1).

**Figure 1.3** Cells of wild-type *Chl. tepidum* (black arrow) and polysulfide granules deposited extracellularly while cells are using sulfide as an electron donor (white arrow). The sulfur (S$_0$) in these granules will itself be used as an electron donor when the sulfide is depleted.

**Figure 1.4** Cross-section of wild-type *Chl. tepidum*; chlorosomes are clearly visible as electron-transparent (light-colored) compartments underlying the cell membrane. Micrograph courtesy of Niels-Ulrik Frigaard.

**Figure 1.5** Model of the *Chl. tepidum* chlorosome. BChl c aggregates in the interior are illustrated both as rods and as lamellae; currently, both models are supported by microscopic and spectroscopic data. BChl c is found only inside the chlorosome, BChl a is found associated with CsmA in the chlorosome baseplate, with the FMO protein between the chlorosome and the cytoplasmic-membrane-bound reaction centers, and in the reaction centers. Chl a is found only in the reaction centers. Carotenoids are associated with BChl c in the interior of the chlorosome, with all of the BChl a-containing pigment-protein complexes including the reaction centers, and in the cytoplasmic membrane. Model diagram courtesy of Hui Li and Niels-Ulrik Frigaard.

**Figure 1.6** Structures of BChls c, d, and e. The C$_3^1$ hydroxyl chelates the Mg in an adjacent BChl molecule and is responsible for the aggregation of the BChls.

**Figure 1.7** Model of the GSB reaction center. The electron comes from heme$_{551}$ in PscC. Cofactors: P840, the primary donor, is BChl a; A$_0$, Chl a; A$_1$, menaquinone-7; F$_x$, F$_A$, and F$_B$ are all [4Fe-4S] clusters. Adapted from Hauska *et al.*, 2001.

**Figure 1.8** Structure of BChl a

**Figure 1.9** Structure of Chl a

**Figure 1.10** Structures of carotenoids found in *Chl. tepidum*. A. chlorobactene. B. γ-carotene. C. 1’,2’-dihydrochlorobactene. D. 1’,2’-dihydro-γ-carotene. E. hydroxychlorobactene glucoside laurate.
Figure 1.11 Structures of some carotenoids found in *Chl. phaeobacteroides* strain CL1401. A. isorenieratene. B. β-isorenieratene. C. phytoene, which accumulates only when the cells are treated with the inhibitor HBP.
References


57. **Hauska, G., T. Schoedl, H. Remigy and G. Tsio**

58. **Heising, S., L. Richter, W. Lud**


60. **Hildebrandt, P., H. Tamiaki, A.R. Holzwarth, and K. Schaff**


64. **Hugler, M., H. Huber, S.J. Molyneaux, C. Vetr**


Figure 1.1

- **Gloeobacter violaceus**
- **Synechococcus sp. PCC 6301**
- **Prochlorococcus marinus MED4**
- **Crocosphaera watsonii**
- **Sphingopyxis alaskensis**
- **Erythrobacter litoralis**
- **Nitrobacter winogradskyi**
- **Rhodopseudomonas palustris**
- **Bradyrhizobium japonicum**
- **Agrobacterium tumefaciens**
- **Rhodobacter sphaeroides**
- **Rhodobacter capsulatus**
- **Escherichia coli strain K12**
- **Shewanella oneidensis**
- **Idiomarina loihiensis**
- **Dechloromonas aromatica**
- **Borrelia bronchiseptica**
- **Bdellovibrio bacteriovorus**
- **Anaeromyxobacter dehalogenans**
- **Geobacter sulfurreducens**
- **Acidobacterium sp. Ellin345**
- **Sollbacter usitatus**
- **Chloroacidobacterium thermophilum**
- **Helicobacter pylori**
- **Campylobacter jejuni**
- **Caldicellulosiruptor saccharolyticus**
- **Clostridium acetobutylicum**
- **Thermoanaerobacter tengcongensis**
- **Helioseptium modesticaldum**
- **Listeria monocytogenes 4b f2365**
- **Bacillus cereus**
- **Staphylococcus aureus COL**
- **Porphyromonas gingivalis**
- **Cytophaga hutchinsonii**
- **Chlorobium tepidum**
- **Chlorobium phaeobacteroides strain MN1**
- **Brevibacterium linens**
- **Arthrobacter strain f24**
- **Streptomyces coelicolor**
- **Dehalococcoides ethenogenes**
- **Chloroflexus aggregens**
- **Chloroflexus aurantius**
- **Roseiflexus castenholzii**
- **Aquifex aeolicus**
- **Thermotoga maritima**
- **Thermus thermophilus**
- **Deinococcus radiodurans**

- **cyanobacteria**
- **α-proteobacteria**
- **γ-proteobacteria**
- **β-proteobacteria**
- **δ-proteobacteria**
- **acidobacteria**
- **ε-proteobacteria**
- **low-GC Gram-positive bacteria**
- **CFB clade**
- **green sulfur bacteria**
- **high-GC Gram-positive bacteria**
- **Chloroflexi**
- **Aquificales**
- **Thermotogales**
Figure 1.3
Figure 1.4
Figure 1.5
Figure 1.6

A.

B.

C.
Figure 1.7
Figure 1.8
Figure 1.9

\[
\Delta_{2,6}\text{-phytadienyl}
\]
Figure 1.10

A.

B.

C.

D.

E.
Figure 1.11

A. \[ \begin{align*} & \text{Structure A} \\ & \text{Diagram of A} \end{align*} \]

B. \[ \begin{align*} & \text{Structure B} \\ & \text{Diagram of B} \end{align*} \]

C. \[ \begin{align*} & \text{Structure C} \\ & \text{Diagram of C} \end{align*} \]
CHAPTER 2

A NEW CLASS OF CAROTENOID CYCLASES

Publications:


Abstract

The genome of Chlorobium tepidum, like the genomes of several cyanobacteria, does not encode a homolog of any previously identified lycopene cyclases. To identify which of the 973 hypothetical proteins in the Chl. tepidum genome encodes a lycopene cyclase, phylogenetic profiling was used to identify candidate genes. A single open reading frame, CT0456, was identified as the gene most likely to encode the lycopene cyclase, and its function was confirmed in a complementation assay in E. coli. All chlorobactene-producing green sulfur bacteria have one ortholog of this gene, now named cruA, and all isorenieratene-producing species of GSB have two homologs, cruA and cruB. Those cyanobacteria whose genomes do not encode crtL-, crtY-, or crtYcYd-type lycopene cyclases have not one but two homologs of cruA. In addition, several plants which have CrtL-type lycopene cyclases also have one homolog of cruA in their genomes. Thus, identification of the Chl. tepidum lycopene cyclase has identified a new class of lycopene cyclases distributed among photosynthetic organisms.
2.1 Introduction
Colored carotenoids are found in nearly all photosynthetic organisms and in a variety of non-photosynthetic ones, and play roles in light-harvesting, photoprotection, structural maintenance of pigment-protein complexes (19), and membrane structure and fluidity (66). The first colored intermediate in many carotenogenic pathways is lycopene, for which two biosynthetic pathways exist. In the plant and cyanobacterial pathway, two enzymes (CrtP and CrtQ in cyanobacteria, PDS and ZDS in plants) successively desaturate phytoene to 7, 9, 7’, 9’ tetra-cis-lycopene (Figure 2.1A; 1, 7, 10, 11, 20, 41, 42, 58), and an isomerase, CrtH (CRTISO in plants), converts cis-lycopene to all-trans-lycopene (Figure 2.1A; 3, 12, 31, 43, 49). In contrast, the CrtI protein in proteobacteria performs three or four successive desaturations to convert phytoene directly to neurosporene or all-trans-lycopene (Figure 2.1B; 21, 26, 41, 54).

The cyclization of lycopene to γ, β, or ε-carotene is an important branching point in the carotenoid biosynthetic pathways in bacteria, plants, and fungi (3, 60). In fact, inhibition of lycopene cyclization has been shown to cause rapid degradation of photosystem II in plants and algae (18) as well as growth defects and a reduction in efficiency of energy transfer in green sulfur bacteria (GSB) (2, 53). Two types of end groups are commonly produced by the cyclization of lycopene: ε-ionone rings, which have a double bond between the C-4 and C-5 positions, and β-ionone rings, which have a double bond between the C-5 and C-6 positions (Figure 2.2). ε-rings are found on lutein, which has one β- and one ε- ring, as well as the rare carotenoid ε-carotene, synthesized in lettuce (13). The monocyclic γ-carotene is a precursor to the myxols in cyanobacteria and plants (46, 64), and is an intermediate in chlorobactene biosynthesis in green-colored GSB (23, 63). The dicyclic β-carotene is a major component of the photosynthetic reaction center complexes in cyanobacteria and plants (25, 32), and β-carotene can be modified to isorenieratene in brown-colored GSB and in some actinomycetes (29, 38, 40).

Three classes of lycopene cyclases have previously been identified in bacteria: the CrtY-type β-cyclases that are found in many carotenogenic proteobacteria (e.g. 44, 45), Streptomyces spp. (38), and the Chlorflexi; the CrtL family, which includes the β- and ε-cyclases in some cyanobacteria and plants (15, 61); and the heterodimeric cyclases of
some Gram-positive bacteria (36, 67), which are related to the lycopene cyclases of the archaea Halobacterium salinarum and Sulfolobus solfataricus (28, 50). These three classes are distantly related to each other and share a few conserved motifs; the first two types include an N-terminal flavin-binding domain that appears to be missing in the third (37). Enzymes that produce monocyclic carotenoids are present in both the CrtY- and CrtL-type classes (13, 37, 62, 64), and there are no obvious sequence differences between mono- and di-cyclases (13).

On the basis of sequence similarity and targeted gene inactivation, most of the chlorobactene biosynthetic pathway in Chlorobium (Chl.) tepidum was previously elucidated (23; see Appendix A). Comparison of the sequences of the crt genes in Chl. tepidum to those in other carotenogenic organisms showed that the carotenoid biosynthetic pathway in Chl. tepidum is more similar to the pathway found in plants and cyanobacteria than to the pathway found in other species of bacteria (23; see Appendix A). The primary carotenoid in green-colored green sulfur bacteria such as Chl. tepidum is chlorobactene, which has one cyclic end group: a lycopene cyclase must exist in the genome of Chl. tepidum. However, no homologs to crtL, crtY, or the heterodimeric cyclases were detected in its genome (17). Additionally, the completed genomes of Gloeobacter violaceus (48), Synechocystis sp. PCC 6803 (34), Synechococcus sp. PCC 7002 (Zhao, Li, and Bryant, personal communication), and Nostoc sp. PCC 7120 (35) also lack any gene homologous to any known lycopene cyclase. Overall, the carotenoid biosynthesis pathways of these organisms are very similar to that of Chl. tepidum, and the number of genes shared by GSB and cyanobacteria is low, thus making phylogenetic profiling a practical way to identify candidate lycopene cyclases.

Phylogenetic profiling compares the presence or absence of a gene in a genome with the presence or absence of a specific trait in the organism (51); genes found only in those organisms which also have a particular phenotype are likely to affect that trait. It is frequently used to predict the functions of conserved hypothetical proteins (27, 51, 56), and the validity of the predictions in several systems has been verified either genetically or biochemically. In eukaryotic systems, phylogenetic profiling has been used to identify probable C9-methyltransferases in the sphingolipid biosynthesis pathway, and the gene functions were confirmed by inactivation and overexpression of the gene in yeast (65).
prokaryotic systems, where ORF prediction is more straightforward and more genomes are available, phylogenetic profiling in conjunction with inactivation of the candidate genes and heterologous expression has been used to confirm the roles of conserved hypothetical proteins (CHPs) in flagellar biosynthesis in *Bacillus subtilis* (39) and isoprenoid biosynthesis in the nonmevalonate pathway in cyanobacteria and *E. coli* (14).

In addition, the herbicide 2-(4-methylphenoxy)triethylamine hydrochloride, which has been used to inhibit lycopene cyclization in other organisms (4, 18, 52), successfully inhibits the same step in GSB (2, 23). It seemed probable that the lycopene cyclase in green sulfur bacteria would share at least some functional domains with other types of lycopene cyclases, most likely the putative flavin-binding domain. The combination of phylogenetic profiling and sequence comparison was used to reduce the number of candidate lycopene cyclases from over 900, the number of conserved hypothetical proteins predicted in the *Chl. tepidum* genome (17), to one.
2.2 Methods

2.2.1 Bacterial strains and growth conditions. Wild-type *Chl. tepidum* is the plating strain WT2321 as described in Wahlund and Madigan (68), which is derived from strain TLS (ATCC 49652; 69). Cells were grown in liquid CL medium or solid CP medium at 42°C. When appropriate, streptomycin and spectinomycin were added to the CP to final concentrations of 150 and 300 µg ml\(^{-1}\) or to the CL to final concentrations of 37.5 and 75 µg ml\(^{-1}\), respectively. Transformation of this strain was performed as previously described (22). *Chl. tepidum* was manipulated in a Coy anaerobic chamber (Coy, Grass Lakes, MI) with an atmosphere of 10% H\(_2\) and 10% CO\(_2\) balanced with N\(_2\). Routine recombinant DNA procedures were performed in electrocompetent DH10B. When preparing constructs with the *aadA* cassette, the strain DH5\(\alpha\) was used instead. When grown for pigment analysis, *E. coli* strains were grown in minimal M9 medium supplemented with glucose (57). When appropriate, *E. coli* strains were grown in or on media amended with 100 µg ml\(^{-1}\) ampicillin, 25 µg ml\(^{-1}\) chloramphenicol, 15 µg ml\(^{-1}\) kanamycin, or 100 µg ml\(^{-1}\) spectinomycin and 50 µg ml\(^{-1}\) streptomycin. Concentrations were not changed when several antibiotics were used together.

2.2.2. Phylogenetic profiling. All predicted proteins of *Chl. tepidum* were searched using blastp against all proteins of completed bacterial genomes as described (17). *Chl. tepidum* proteins were then clustered according to the patterns of species in which homologs were found (51).

2.2.3 Preparation of genomic library and identification of gene. *Chl. tepidum* genomic DNA was digested with the enzyme *Sau3AI* (1.25 \(\times\) \(10^2\) U µg\(^{-1}\) DNA). Fragments of 2.5 to 5 kb were excised from an agarose gel, purified with the PerfectPrep Gel Cleanup Kit (Eppendorf, Hamburg, Germany), and ligated into pUC19 that had been digested with *BamHI*. The library was introduced by electroporation into *E. coli* strain DH10B carrying the pAC-LYC plasmid (15). Transformants were plated on LB plates containing 25 µg chloramphenicol ml\(^{-1}\) and 100 µg ampicillin ml\(^{-1}\) and visually screened for development of orange color. The plasmids from the single orange transformant (plasmid pCTLY, Fig. 2.1A) were extracted after growth for 8 h in liquid media, and the ends of the cloned *Chl.*
tepidum DNA insert were sequenced using the M13U and M13R primers at the Nucleic Acid Facility at Penn State University.

2.2.4 Construct for inactivation of cruA (CT0456) in Chl. tepidum. The aadA cassette from plasmid pSRA2 (22) conferring resistance to streptomycin and spectinomycin was inserted into plasmid pCTLY at the MscI site (Figure 2.1B). This plasmid was linearized with Scal and the linearized plasmid was used to transform wild-type Chl. tepidum as previously described (22).

2.2.5 Pigment analysis. Pigments were extracted from cell pellets by sonication in 0.4 ml acetone:methanol (7:2 v:v). The extract was centrifuged to remove cell debris, and the supertatant was filtered through a 0.2 µm syringe filter (Whatman) prior to injection into the HPLC. Pigments were separated by HPLC (Agilent Model 1100 equipped with a diode array detector [model G1315B] and controlled with Agilent ChemStation software; Agilent Technologies, Palo Alto, CA) on a 25 cm by 4.6 mm Discovery 5 µm C18 column (Supelco, Bellefonte, PA). The solvent system for analysis of all strains was modified from the previously described system (23, 24). Solvent B was as previously published (23, 24), but solvent A consisted of water:methanol:acetonitrile (62.5:21:16.5) containing 10 mM ammonium acetate. For Chl. tepidum strains, at the time of sample injection, the solvent contained 20% B, which was linearly increased to 70% B over 10 min., linearly increased to 100% B over the next 30 min., then held at 100% B for 10 min. before returning to 20% B over 10 min. For analysis of E. coli strains, the solvent contained 20% B at injection, which was linearly increased to 100% B over 30 min and then held constant at 100% B for 5 min. The flow rate was 0.750 ml min$^{-1}$ until the proportion of solvent B reached 30%, at which time the flow rate was increased to 1 ml min$^{-1}$. Pigments were identified on the basis of their retention time, absorption spectra, and mass.
2.3 Results

2.3.1 Phylogenetic profiling. Several cyanobacteria that produce γ- or β-carotene and derivatives thereof lack homologs of known lycopene cyclases. Given the overall similarity of the carotenoid biosynthetic pathways of cyanobacteria and *Chl. tepidum* (23; see Appendix A), we hypothesized that the lycopene cyclase from *Chl. tepidum* would also be found in these cyanobacteria. Because the *Prochlorococcus* species whose genomes have been sequenced each encode two CrtL homologs (55), these species would not be expected to encode homologs of the lycopene cyclase in *Chl. tepidum*. The completely sequenced genomes of *Chl. tepidum*, *Synechocystis* sp. PCC 6803, *Nostoc* sp. PCC 7120, *Thermosynechococcus elongatus* and *Gloeobacter violaceus* were systematically searched using blastp for conserved hypothetical proteins in *Chl. tepidum* that were not found in the genomes of *Chloroflexus* (*Cfx.*.) *aurantiacus* or any *Prochlorococcus* spp. This phylogenetic profiling analysis identified three *Chl. tepidum* ORFs: *CT0456*, *CT1196*, and *CT1748* (Fig. 2.2). The best-characterized motif shared by most lycopene cyclases is an FAD-binding motif (37), and only ORF *CT0456* had such a motif, making it the gene most likely to encode a lycopene cyclase.

2.3.2 Identification of Lycopene Cyclase in *Chl. tepidum*. *E. coli* cells harboring plasmid pAC-LYC produce pink colonies containing lycopene (Figure 2.3A), and these cells were transformed with a genomic library from *Chl. tepidum*. Any cells expressing the *Chl. tepidum* lycopene cyclase should convert lycopene into γ-carotene and thus form orange colonies. One orange transformant was obtained, and the orange-colored cells produced γ-carotene and small amounts of β-carotene and torulene (3’4'-didehydro-γ-carotene) (Fig. 2.3B). The library plasmid from the orange colony, denoted pCTLY (Figure 2.4), had an insert of approximately 2.5 kb, which encoded one partial gene (DNA ligase, ORF *CT0457*) and one complete gene (*CT0456*). *CT0456* encodes a conserved hypothetical protein that has now been designated CruA.

2.3.3 Inactivation of Lycopene Cyclase in *Chl. tepidum*. After transformation with linearized pCTLY::*aadA* (Figure 2.5A), *Chl. tepidum* transformants resistant to streptomycin and spectinomycin appeared on plates within a week. After several passages on solid selective medium, transformants were grown in liquid media for pigment and DNA
analysis. PCR amplification of the CT0456 locus confirmed that the mutant and wild-type alleles had completely segregated (Fig. 2.5B). The pigments produced by wild-type Chl. tepidum have been identified by HPLC and LC-MS as mainly chlorobactene, some γ-carotene, and small amounts of glycosylated and acylated derivatives of these two compounds (23, 63; Figure 2.6). Similar analysis of the cruA::aadA mutant found that this strain produced mainly lycopene (Fig. 2.6, peak 5) as well as smaller amounts of lycopene precursors and acyl-glucoside esters of hydroxylcopene. These data demonstrate that the cruA::aadA mutant is unable to produce carotenoids with cyclic end groups.

2.3.4 Phylogenetic Analyses of Lycopene Cyclases. A single cruA gene was identified in the draft genomes of Chl. chlorochromatii, Pld. luteolum, Chl. vibrioforme f. thiosulfatophilum, Chl. limicola, and Prosthecochloris aestuarii, the GSB that produce chlorobactene. The amino acid sequences of these predicted proteins are 60 to 67% identical and 77 to 80% similar to Chl. tepidum CruA. In contrast, Chl. phaeobacteroides and Pld. phaeoclathratiforme are brown-colored GSB that produce isorenieratene; each of their genomes encodes two CruA homologs. In both species, one homolog is 64% identical and 70 to 80% similar, and the other is 54% identical and 71% similar to Chl. tepidum CruA. This second homolog in brown species has been named cruB.

A blastp search using the deduced amino acid sequence of the C. tepidum CruA identified only two groups of homologous proteins in the NCBI database (http://www.ncbi.nlm.gov/BLAST), all of which are conserved hypothetical proteins in the FixC superfamily of dehydrogenases. FixC is predicted to be a membrane-integral flavoprotein involved in electron transfer to nitrogenase (16). The first group of homologous proteins includes CruA, CruB, and CruA orthologs from several cyanobacterial species. None of these cyanobacteria, which include Synechococcus sp. PCC 7002, Synechococcus sp. WH8501, Synechocystis sp. PCC 6803, Gloeobacter violaceus, Anaabaena sp. PCC 7120, and Nostoc punctiforme, has a CrtL-type lycopene cyclase, but all encode two CruA homologs. The second group includes paralogs of CruA, denoted CruP, that are found in the same cyanobacterial strains and in plants. The only cyanobacteria that have members of both the CrtL and CruA families are the closely related Synechococcus sp. strains PCC 7942 and PCC 6301. The two groups of CruA homologs
form a clade that is distinct from the CrtY-, CrtL-, and heterodimeric-type lycopene cyclases (Figure 2.7).

2.4 Discussion

Identification of CruA as the sole lycopene cyclase in green-colored GSB has enabled identification of a putative isorenieratene-specific cyclase, CruB, in brown-colored GSB as well as two homologous carotenoid cyclases in cyanobacteria, CruA and CruP (Maresca, Graham, et al., in preparation). Additionally, all genes encoding enzymes required for the synthesis of chlorobactene in Chl. tepidum have now been identified (Fig. 2.8; see chapter 4). The other genes in this pathway were identified on the basis of their similarity to genes involved in carotenoid biosynthesis in other organisms, and their functions were confirmed by targeted gene inactivation (23). However, this approach missed the lycopene cyclase in Chl. tepidum as well as the carotenoid glycosyltransferase and acyltransferase (Chapter 4), and consequently phylogenetic profiling was used to identify ORFs potentially encoding the lycopene cyclase. Because the number of genes shared between GSB and cyanobacteria is low (17) and because most known lycopene cyclases also share an FAD-binding motif (37), this method correctly identified CT0456 as the most likely candidate. At the same time, a genomic library from Chl. tepidum was used to complement a lycopene-producing strain of E. coli, making it a γ-carotene-producing strain. The gene encoding the lycopene cyclase is ORF CT0456, now named cruA.

Orthologs of CruA are found in all of the cyanobacterial genomes that lack CrtL-type lycopene cyclases, and identification of the carotenoid cyclases in these strains resolves a question that has existed since the genome of Synechocystis sp. PCC 6803 was completed and was found not to encode any homologs to CrtL, CrtY, or heterodimeric cyclases (34). Several plant genomes also have CruA homologs; interestingly, of the over 500 microbial genomes that are either completed or available in draft form in the JGI, TIGR, and NCBI databases, only organisms with Type I reaction centers have CruA-type lycopene cyclases, with only one exception, the nonphototrophic Herpetosiphon aurantiacus. It is possible that the genes encoding type I reaction centers and carotenoids were inherited from a common ancestor.
The **cruA** mutant of *Chl. tepidum* accumulates lycopene, which indicates that the substrate of CruA is lycopene (also see Chapter 3). Because β-carotene and its derivatives isorenieratene and β-isorenieratene have not been observed in *Chl. tepidum* (23, 63), CruA must strictly be a monocyclase in *Chl. tepidum*, despite the fact that in *E. coli*, CruA has some dicyclase activity.

Although CruA, like CrtY and CrtL, has a putative N-terminal flavin-binding domain (35), none of these proteins has yet been shown to bind flavins. Flavoproteins frequently catalyze oxidation-reduction reactions; however, the conversion of lycopene to γ- or β-carotene is an isomerization, and there is no net redox change. Although the *Erwinia uredovora* CrtY requires NADPH for catalysis, NADPH does not contribute any hydrogens to the final product of the reaction (30, 59). There are other flavoproteins that catalyze reactions in which there is no net redox change (see reference [9] for examples); however, CruA does not resemble any of these proteins in sequence.

It is possible that its catalytic function is most similar to that of the Type II isopentenyl diphosphate (IPP)/dimethylallyl diphosphate (DMAPP) isomerase, which converts IPP to DMAPP by isomerizing a double bond (5, 33). Two types of IPP isomerases have been described, and the Type II enzyme, like CrtY, requires a reduced flavin and NADPH for activity (5, 6, 33). Both Type I and Type II IPP/DMAPP isomerases appear to catalyze isomerization by stereoselective protonation and deprotonation of the substrate (6, 70; Figure 2.10). In this case, although there is no net oxidation or reduction, both reduction and oxidation of the substrate occur during the reaction, and the flavin may have a role in these half-reactions. Because the cyclization of lycopene is also an isomerization of an isoprene unit (see Figure 2.2), the flavins in lycopene cyclases may have a function similar to that of the flavin in Type II IPP/DMAPP isomerase.

CruA is the first-characterized member of what appears to be a family of carotenoid cyclases that has homologs only in phototrophic organisms. This distribution raises interesting questions about why, when horizontal transfer of carotenoid biosynthetic genes appears to have occurred frequently between completely unrelated microorganisms, this gene has not. The genes encoding phytoene synthases, carotenoid desaturases, and carotenoid glycosyltransferases can be found not only in photosynthetic bacteria but also in *Deinococcus radiodurans*, the planctomycete *Rhodopirellula baltica*,

48
and actinomycetes (see Chapter 4). However, $\textit{cruA}$ is found only in cyanobacteria, plants, GSB, and the FAP $\textit{Herpetosiphon aurantiacus}$. CruA may require another protein, also found exclusively in cyanobacteria, plants, and GSB, for activity. Perhaps it requires a cofactor not synthesized by many bacteria. It is also possible that, although it can act on the all-$\textit{trans}$ lycopene made in $\textit{E. coli}$, its substrate $\textit{in vivo}$ is instead an intermediate specific to the CrtPQH pathway.
2.5 References


Chapter 2 Figure Legends.

**Figure 2.1** Synthesis of lycopene. **A.** Cyanobacterial pathway for production of lycopene from phytoene. The intermediates in this pathway are *cis*-carotenoids, which are converted to all-*trans*-lycopene by the isomerase CrtH. **B.** Proteobacterial pathway for production of lycopene from phytoene. CrtI converts phytoene directly to all-*trans*-lycopene or neurosporene, depending on the species of bacterium.

**Figure 2.2** Cyclization of lycopene. *ε*-cyclases such as those found in plants (LCYe) and *Prochlorococcus* (CrtLe) catalyze isomerization of a *ψ*-end to an *ε*-ring. *β*-cyclases such as those found in plants (LCYb), cyanobacteria (CrtL or CrtLb), Gram-positive bacteria (CrtYcYd), and many other kinds of bacteria (CrtY or CrtL) catalyze isomerization of a *ψ*-end to a *β*-ring. The unidentified lycopene cyclases in GSB and some cyanobacteria are *β*-cyclases.

**Figure 2.3** Phylogenetic profiles of selected *Chl. tepidum* ORFs. Dark boxes indicate presence of an ORF in a specific genome; light boxes indicate absence. ORFs marked with open circles are enzymes in bacteriochlorophyll biosynthesis; those marked with filled circles are known to be involved in carotenogenesis.

**Figure 2.4.A.** Elution profiles of carotenoids produced by *Escherichia coli* with plasmids pAC-LYC and pUC19 (upper panel) or pAC-LYC and pCTLY (lower panel). Peaks: 1, lycopene; 2, torulene; 3, γ-carotene; 4, β-carotene. **B.** Absorption spectra of lycopene (peak 1, upper panel) or γ-carotene (peak 3, lower panel).

**Figure 2.5** Plasmid pCTLY encoding part of *CT0457*, predicted to encode DNA ligase, and the complete ORF *CT0456*, the lycopene cyclase, now named *cruA*.

**Figure 2.6.** Construction and verification of *CT0456* inactivation mutant. **A.** Construct for inactivation. The cassette encoding *aadA*, conferring resistance to streptomycin and spectinomycin, was inserted into the MscI site in plasmid pCTLY. **B.** PCR verification of segregation of the mutant. Lanes labeled M contain a DNA size marker, Ladder I from GeneChoice. Templates in lanes 1, 2, and 3 are from wild-type cells and two isolates of the *CT0456* deletion mutant. Primers amplifying *cruA* (*CT0456*) clearly show a 1.1 kb insertion in the *CT0456* locus in the two mutant isolates; primers amplifying the *aadA* gene demonstrate presence of that gene only in the mutants.

**Figure 2.7.** Elution profiles of wild-type *Chl. tepidum* (dark line) and *CT0456* mutant (gray line). Peak 1, hydroxychlorobactene glucoside laurate; peak 2, chlorobactene, peak 3, 1’, 2’-didehydrochlorobactene; peak 4, γ-carotene. The mutants accumulates only lycopene (peak 5), and small quantities of compounds with the same chromophore.
Figure 2.8 Neighbor-joining tree showing phylogeny of known lycopene cyclases. Bold lines indicate proteins whose function has been confirmed either genetically or biochemically.

Figure 2.9 Biosynthetic pathway of chlorobactene in *Chl. tepidum*. All genes in this pathway have been inactivated and their functions confirmed.

Figure 2.10 Proposed isomerization reaction mechanism catalyzed by isopentenyl diphosphate (IPP) isomerases. The C3-C4 double bond of IPP is first protonated to produce a carbocation intermediate, then a proton is eliminated to yield a C2-C3 double bond. This figure is adapted from the work of Wouters *et al.* (2003) and Barkley *et al.* (2004).
Table 2.1 Primers used in this work.

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Purpose</th>
<th>Primer sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>CT0456-F4</td>
<td>Confirming segregation</td>
<td>ATC GCA TCC ACA AGC ACC TG</td>
</tr>
<tr>
<td>CT0456-R4</td>
<td>Confirming segregation</td>
<td>GGT GAT GTC GAA AAA GTA CGG G</td>
</tr>
<tr>
<td>M13U</td>
<td>Sequencing plasmid insert</td>
<td>TGT AAA ACG ACG GCC AGT</td>
</tr>
<tr>
<td>M13R</td>
<td>Sequencing plasmid insert</td>
<td>CAG GAA ACA GCT ATG ACC</td>
</tr>
</tbody>
</table>
Table 2.2 Procedures for HPLC analysis of bacterial pigments.

<table>
<thead>
<tr>
<th>Time (min)</th>
<th>% B</th>
<th>Flow rate (ml min⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chl. tepidum</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>20</td>
<td>1.00</td>
</tr>
<tr>
<td>10</td>
<td>70</td>
<td>1.00</td>
</tr>
<tr>
<td>40</td>
<td>100</td>
<td>1.00</td>
</tr>
<tr>
<td>50</td>
<td>100</td>
<td>1.00</td>
</tr>
<tr>
<td>60</td>
<td>20</td>
<td>1.00</td>
</tr>
<tr>
<td>E. coli</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>20</td>
<td>0.75</td>
</tr>
<tr>
<td>3</td>
<td>30</td>
<td>1.00</td>
</tr>
<tr>
<td>30</td>
<td>100</td>
<td>1.00</td>
</tr>
<tr>
<td>35</td>
<td>100</td>
<td>1.00</td>
</tr>
<tr>
<td>40</td>
<td>20</td>
<td>0.75</td>
</tr>
</tbody>
</table>
Figure 2.1

A. 15, 15' cis-phytoene
    \[\text{CrtP} \]
    9, 9' di-cis-ζ-carotene
    \[\text{CrtQ} \]
    7,9,7',9' tetra-cis-lycopene
    \[\text{CrtH} \]
    all-trans- lycopene

B. 15, 15' cis-phytoene
    \[\text{CrtL} \]
    \[\text{CrtL} \]
    neurosporene
    all-trans- lycopene
Figure 2.2

\[
\begin{align*}
\text{ψ end group} & \\
\text{ε cyclase} & \quad \text{β cyclase} \\
\text{ε ring} & \\
\text{β ring}
\end{align*}
\]
Figure 2.3
Figure 2.4
Figure 2.5
Figure 2.6
Figure 2.7
Figure 2.8
Figure 2.9

Geranylgeranyl-PP → CrtB → 15-cis-phytoene → CrtP → 9,9’-di-cis-ζ-carotene → CrtQ → 7,9,7’,9’-tetra-cis-lycopene → CrtH → all-trans-lycopene → CruA → γ-carotene → CrtU → chlorobactene
Figure 2.10

![Chemical Structures]

- Isopentenyl diphosphate (IPP)
- Dimethylallyl diphosphate (DMAAPP)
CHAPTER 3

HETEROLOGOUS EXPRESSION OF CRUA-TYPE CAROTENOID CYCLASES
AND PROPOSED BIOSYNTHETIC PATHWAY FOR ISORENIERATENE
Abstract

The recently identified CruA-type of lycopene cyclases is found exclusively in photosynthetic bacteria. Homologs of the \textit{Chl. tepidum} CruA are found in all available genomes of green sulfur bacteria (GSB) and two homologs are found in the genomes of brown-colored species of GSB. Heterologous expression of \textit{cruA} from \textit{Chl. tepidum} and \textit{Chl. phaeobacteroides} strain 266\textsuperscript{T} suggests that CruA is a lycopene monocyclase in all GSB, and CruB is a \(\gamma\)-carotene cyclase in brown species. Inhibition of lycopene cyclization in the brown-colored strain \textit{Pelodictyon phaeum} leads to accumulation of lycopene and neurosporene, but the strain \textit{Chl. phaeobacteroides} strain 1549 accumulates neurosporene and neurosporene derivatives both under normal growth conditions and when cyclization is inhibited. In the proposed pathway for isorenieratene biosynthesis, lycopene is the substrate for CruA and \(\gamma\)-carotene is the substrate for CruB; however, an alternate pathway is proposed for carotenogenesis in \textit{Chl. phaeobacteroides} strain 1549, which may have a less-active variant of the \(\zeta\)-carotene/neurosporene desaturase CrtQ.
3.1 Introduction

The pigmentation differences between green-colored and brown-colored green sulfur bacteria (GSB) are not merely cosmetic: species with different pigmentation inhabit different parts of the photic zone. Green-colored species, which make BChl $c$ or $d$ as their primary antenna BChl (27, see chapter 5) and chlorobactene (Figure 3.1) as their main carotenoid, tend to be found in the upper part of anaerobic photic zones, where the light is relatively enriched in longer wavelengths (16, 20, 21, 31, 32). Brown-colored species, which make BChl $e$ (14, see chapter 6) and the dicyclic carotenoid isorenieratene (Figure 3.1), are usually found deeper in the water column, where the light intensity is lower and the light is enriched in green wavelengths (5, 31, 32).

This small change in carotenoid content, from monocyclic carotenoids to dicyclic, has an immediate consequence. The absorption spectra of chlorobactene and BChl $e$ overlap, but the absorption spectra of $\beta$-carotene and its derivatives are blue-shifted relative to chlorobactene, which broadens the in vivo absorption spectrum in the green wavelengths. The energy transfer efficiency from isorenieratene to BChl $e$ in chlorosomes has been measured to be as high as 70% (24). Thus it seems that isorenieratene may be an important accessory antenna pigment in the brown species, which live at lower light intensities than the green species (5, 21, 31, 32) and require larger antennas with broader absorption peaks.

Three of the GSB species whose genomes have been sequenced are brown-colored GSB, and a BLAST search (2) against the genomes of Chl. phaeobacteroides strain 266$^T$, Chl. phaeobacteroides strain MN-1, and Pelodictyon (Pld.) phaeoclathratiforme strain BU-1 using CruA from Chl. tepidum as a query sequence identified two CruA homologs in each of these genomes. One homolog groups with the cruA sequences of the green species; the other, denoted cruB, groups separately within the cruA clade (Figure 3.2). If CruA is a lycopene monocyclase in all species, then CruB may be a $\gamma$-carotene cyclase. Alternatively, CruB may be a lycopene dicyclase specific to isorenieratene biosynthesis, while CruA is specific to chlorobactene biosynthesis; isorenieratene-producing species synthesize a small amount of glycosylated and acylated OH-chlorobactene (15). A third possibility is that the product of CruB is not a compound
with a β-ionone ring at all, but instead some other type of ring. Both the CrtL and heterodimeric (CrtYcYd) families of lycopene cyclases have members that synthesize ε rather than β rings (8, 15, 23), and it may be that a minor carotenoid component of brown-colored GSB requires a different type of cyclase. Because it is not yet possible to inactivate genes in any brown-colored GSB species (see Appendix B), these hypotheses were tested using heterologous expression in *Chl. tepidum* and *E. coli*.
3.2 Methods

3.2.1 Strains and growth conditions. *Chl. tepidum* strains are as described in Chapter 2. *Chl. phaeobacteroides* strain 1549 and *Pelodictyon (Pld.) phaeum* are brown-colored species of GSB capable of using thiosulfate as an electron donor and were obtained from Dr. John Ormerod, University of Oslo and Dr. Robert Blankenship, Arizona State University, respectively. *Chl. phaeobacteroides* strain 1549 and *Chl. tepidum* were grown in CL medium (12), and *Pld. phaeum* was grown in CL medium with additional 3% NaCl. *Chl. tepidum* was grown at 40-46°C, and *Chl. phaeobacteroides* and *Pld. phaeum* were grown at 28-30°C. *Escherichia coli* strain BL21 (DE3) with plasmid pAC-LYC (encoding genes for synthesis of lycopene; 10) or pAC-NEUR (encoding genes for synthesis of neurosporene; 10) was used for complementation experiments and was grown at 28-30°C either on LB plates or in M9 medium (25) supplemented with 25 µg ml⁻¹ chloramphenicol, 100 µg ml⁻¹ ampicillin, and/or 15 µg ml⁻¹ kanamycin, as necessary.

3.2.2 Constructs for expression of CruA and CruB in *E. coli*. Because 100% of the available lycopene was converted to cyclic carotenoids by *Chl. tepidum* CruA expressed from its own promoter in the plasmid pCTLY, no other expression construct was made with this gene. The genes *cruA* and *cruB* were amplified from the genomic DNA of *Chl. phaeobacteroides* strain DSM 266ᵀ using primer pairs CPL1-F3/CPL1-R1 and CPL2-FN/CP2404-KR (Table 3.1). The PCR products were subcloned into vector pCR2.1-TOPO using the TOPO-TA cloning kit (Invitrogen, Carlsbad, CA, Catalog # K4500-40), to produce plasmids pTA-CPL1 and pTA-KcruB. The fragments including *cruA* and *cruB* were then excised from those plasmids using the enzymes *Nde*I and *Bam*HI for *cruA* and *Nco*I and *Kpn*I for *cruB*. The fragments were inserted into vectors pET16b (Novagen, La Jolla, CA, Catalog #69662-3) and pCOLA-DUET1 (Novagen, La Jolla, CA, Catalog #71406-3), respectively, generating plasmids pCPL1 and pCOLA-B (Figures 3.3A and B).

3.2.3 Assays of CruA and CruB activity in *E. coli*. Combinations of plasmids pUC19, pET16b, pCOLA-DUET1, pCTLY, pCPL1, and pCOLA-B were transformed into electrocompetent BL21 (DE3) cells harboring plasmid pAC-LYC (Table 3.2) or pAC-
NEUR (Table 3.3), and cells were plated on LB medium amended with 25 µg ml⁻¹ chloramphenicol, 15 µg ml⁻¹ kanamycin and 100 µg ml⁻¹ ampicillin. Colonies became pigmented without induction after ~48 h at room temperature. For growth in liquid medium, colonies were picked and grown to stationary phase in M9 medium supplemented with the same antibiotics. Genes expressed in the pCOLA vector were induced by addition of 0.4 mM IPTG to the culture when the optical density (OD) at 600 nm had reached ~0.6.

3.2.4 Expression of cruB in Chl. tepidum. The plasmid pCFT-X, in which the gene encoding the Chl. tepidum C-20 methyltransferase, bchU, is interrupted with the cassette for streptomycin and spectinomycin resistance (19, Chapter 5) was used as a platform for heterologous expression of foreign genes in Chl. tepidum. An NcolI site was inserted at the start codon of bchU to produce plasmid pCFT-XN, and a 6.5 kb gene cluster from Pelodictyon (Pld.) phaeoclathratiforme was inserted between the NcolI site and a downstream KpnI site, generating plasmid pXN-BEBj. The aadA cartridge was then re-inserted at the KpnI site, generating plasmid pXN-BEBj-aadA (Figure 3.4). Plasmid pXN-BEBj-aadA was linearized with AhdI and the linear plasmid was used to transform wild-type Chl. tepidum as previously described (12).

3.2.5 Inhibition of carotenoid cyclization. The bleaching herbicide N,N-diethyl-N-[2-(4-methylphenoxy)ethyl]amine (MPTA) has been shown to specifically inhibit lycopene cyclization in a variety of species, including plants (11, 22, 23), cyanobacteria (7, 10), proteobacteria (26) and green sulfur bacteria (13; see Appendix A). Wild-type Chl. tepidum was grown with 3 µM or 20 µM MPTA at 100 µmol photons m⁻² s⁻¹. Chl. phaeobacteroides strain 1549 and Pla. phaeum were grown at 63 and 10 µmol photons m⁻² s⁻¹ and treated with 20 µM MPTA.

3.2.6 Pigment analysis. Pigment analysis of E. coli and Chl. tepidum by HPLC was as described in Chapter 2. For analysis of carotenoids of brown-colored strains, the HPLC protocol used the solvent system described in Chapter 2 and a gradient starting with 70% B, increasing to 94%B over 10 min., then increasing to 100% B over 32 min. The solvent was held at 100% B for 6 minutes, then returned to 70% B.

The identification of the carotenoids produced in E. coli was made based primarily on elution time and absorption spectra. The ~1.5-minute change in elution time
between authentic neurosporene (Figure 3.6A) and the product of the reaction of CruA from *Chl. phaeobacteroides* and neurosporene (Figure 3.6B) corresponds to the ~1.5 minute change in elution between lycopene and γ-carotene (Figures 3.5A and B). Conversion of γ-carotene to β-carotene also causes a ~1.5-minute shift in elution time, similar to the product of CruA from *Chl. tepidum* or CruB from *Chl. phaeobacteroides* on neurosporene. Comparison with the elution times of carotenoids produced by *Chl. phaeobacteroides* strain 1549 during exponential phase, which synthesizes neurosporene and the cyclized neurosporene derivatives 7,8-dihydro-γ- and β-carotene (12) also indicated the same (data not shown).
3.3 Results

3.3.1 Activity of cruA and cruB on lycopene in E. coli. To confirm the hypothesis that cruA in brown-colored species of GSB also encodes a lycopene monocyclase and that cruB encodes a γ-carotene cyclase, cruA and cruB from the brown-colored species Chl. phaeobacteroides strain 266\textsuperscript{T} were expressed singly and together in the lycopene-producing E. coli strain BL21 (DE3) harboring the plasmid pAC-LYC. This E. coli strain was transformed with expression plasmids pCTLY, pCPL1 or pCOLA-CRUB, which encode cruA from Chl. tepidum, and cruA and cruB from Chl. phaeobacteroides strain 266\textsuperscript{T}, respectively, or with the empty vectors pUC19, pET16b or pCOLA (Table 3.3). The cells with plasmids pAC-LYC and the empty vectors pUC19 or pET16b and pCOLA-DUET1 produced only lycopene (Table 3.2); those with pAC-LYC and pCTLY or pCPL1 with pCOLA-DUET1 converted all available lycopene to γ-carotene (Table 3.3, Fig. 3.5C) and those with pAC-LYC, pET16b, and pCOLA-B accumulated lycopene and a small amount of β-carotene but no γ-carotene (Fig. 3.5D). When either cruA was expressed with cruB, almost all of the available lycopene was converted to β-carotene (Fig. 3.5E).

3.3.2 Activity of cruA and cruB on neurosporene in E. coli. The plasmids pCTLY, and pCPL1, encoding the Chl. tepidum CruA and Chl. phaeobacteroides strain 266\textsuperscript{T} CruA, respectively, were transformed either singly or with pCOLA-B into E. coli BL21(DE3) pAC-NEUR (Table 3.3), which produces neurosporene (Figure 3.6, peak 1). The Chl. phaeobacteroides CruA converts the available neurosporene to 7,8-dihydro-γ-carotene, which has an absorption spectrum identical to that of neurosporene but elutes ~1.5 minutes later (Figure 3.6C, peak 2). CruA from Chl. tepidum converts all of the neurosporene to the dicyclic 7,8-dihydro-β-carotene (Figure 3.6B, peak 3). CruB alone appears to have less activity on neurosporene, though the main product is also 7,8-dihydro-β-carotene. When either CruA is expressed with CruB, the primary product is again 7,8-dihydro-β-carotene.

3.3.3 Expression of cruB in Chl. tepidum. The strain of Chl. tepidum in which bchU had been replaced by a gene cluster from Pld. phaeoclathratiforme encoding a putative
radical-SAM-type protein, a putative short-chain dehydrogenase, a putative aldolase, and cruB as well as the aadA cassette from pSRA2 produced streptomycin/spectinomycin-resistant colonies within a few days. After two passages on solid media, the pigments of this strain were analyzed. Because the gene cluster from Pld. phaeoclathratiforme was inserted into the gene encoding the BChl c C-20 methyltransferase, the transformants produced BChl d rather than BChl c; the absence of any detectable BChl c demonstrated that the insertion was completely segregated. The primary carotenoid produced in this strain was β-isorenieratene rather than chlorobactene (Figure 3.7).

3.3.4 Inhibition of lycopene cyclase activity. When the activity of a specific enzyme is inhibited, the substrate of that enzyme should accumulate. To confirm that the substrate for CruA is the same in vivo in green-colored and brown-colored GSB strains, cyclization was inhibited by treatment with the lycopene cyclase inhibitor MPTA in Chl. tepidum (a green-colored strain) and the two brown-colored strains Chl. phaeobacteroides strain 1549 and Pld. phaeum. Under these conditions, Chl. tepidum accumulated primarily lycopene (Figure 3.8A). Chl. phaeobacteroides strain 1549 grown at 63 and 10 µmol photons m\(^{-2}\) s\(^{-1}\) with 20 µM MPTA accumulated lycopene, neurosporene and 7,8 dihydro-γ-carotene (Figure 3.8B). Pld. phaeum grown under the same conditions as Chl. phaeobacteroides strain 1549 accumulated primarily lycopene, some neurosporene, and a small amount of isorenieratene (Figure 3.8C). When MPTA was added to final concentrations of 3 or 100 µM to E. coli strain DH10B with plasmids pAC-LYC and pCTLY, most of the available lycopene was cyclized, but the amount of β-carotene was reduced slightly (Figure 3.9).
3.4 Discussion

In *Chl. tepidum* and other green-colored, chlorobactene-producing species of GSB, CruA appears to be the only carotenoid cyclase, whereas in the brown-colored, isorenieratene-producing species, there are two carotenoid cyclases, CruA and CruB. In *E. coli* strains with plasmid pACLYC, which synthesize lycopene, CruA from both *Chl. tepidum* and *Chl. phaeobacteroides* strain 266^T is a lycopene monocyclase, while CruB has very little activity on lycopene. When co-expressed with CruA, CruB converts γ-carotene to β-carotene very efficiently. In both the CrtL and CrtY families of lycopene cyclases, both mono- and dicyclases exist, and changing only a few amino acids can change the protein from one to the other (8). However, the orthologs of CruA found in GSB all appear to be monocyclases, and a second protein is required to convert the monocyclic carotenoid to the dicyclic form. The pattern of a single CruA ortholog in green species and an ortholog and a paralog (a gene arising from a gene duplication event) in brown species is consistent not only in the 12 species for which complete genomes are available but also in the culture collection of Jorg Overmann (see Chapter 6).

This pattern is interesting given the reported absolute correlation between presence of BChl e and isorenieratene in brown-colored species of GSB. It is possible that this correlation is driven by BChl content: acquisition of a gene or set of genes required for BChl e biosynthesis creates selective pressure that favors mutations in *cruA* that produce variants that can synthesize isorenieratene. However, if this were the case, all species of GSB regardless of color would have single orthologs of *cruA*, with some sequence variation that enables dicyclase activity. Instead, all of the brown species have acquired the gene necessary for isorenieratene biosynthesis. This raises the additional question of whether acquisition of CruB (and subsequent blue-shifting of the major carotenoid absorption peak) creates strong enough selective pressure that a green isorenieratene-producing species would benefit from the ability to oxidize the C7 position of its antenna BChl. This oxidation is a difficult one to perform anaerobically, and it is unlikely that proteins capable of catalyzing this reaction would have arisen spontaneously several times. Thus, it is more likely that *cruB* is one of a small number of genes that are
easily transferred from one species to another, which would confer both the ability to make isorenieratene and BChl e.

In *E. coli*, CruA from *Chl. tepidum* converts about 30% of the available lycopene to β-carotene. This suggests that CruA can have dicyclase activity, and raises the following interesting question: How does *Chl. tepidum* ensure that CruA always generates an asymmetric product? It is possible that either rapid modification of the ψ-end of γ-carotene by CrtC or rapid transport of chlorobactene (φ, ψ-carotene) into the chlorosome could prevent cyclization of the ψ-end. This is unlikely, however, because less than 10% of the total carotenoids have hydroxylated ψ-ends (13, 30), and at least one step (CrtU-catalyzed ring desaturation and methyltransfer) must occur between the cyclization of lycopene and the transport of the product into the chlorosome.

It is possible, instead, that the substrate for CruA is itself asymmetric. The most likely asymmetric substrates would be neurosporene or a cis-isomer of lycopene (see Figure 3.11 for structures). In this case, CruA would only recognize one end as the substrate, and generate an asymmetric product. If CruA is strictly a cyclase and has no oxidizing capability, the product of neurosporene cyclization would be either 7,8-dihydro-γ-carotene or β-zeacarotene, depending on which end is cyclized *in vivo*, and further oxidation of the 7,8 or 7′,8′ bond would be required by a separate desaturase. CrtQ from cyanobacteria is able to desaturate the 7′,8′ bond of β-zeacarotene to produce γ-carotene (1), and the *Chl. tepidum* CrtQ belongs to the same family of enzymes (13) and might have the same activity. However, in neurosporene-producing *E. coli*, CruA from *Chl. tepidum* cyclizes both ends, producing 7,8 dihydro-β-carotene: although CruA differentiates between the identical ψ-ends of lycopene (Figure 3.5), it does not differentiate between the ψ- and 7,8 dihydro-ψ-ends of neurosporene (Figure 3.6). Additionally, if neurosporene were the substrate, it should accumulate in the *cruA* mutant; however, this mutant accumulates primarily lycopene and lycopene derivatives (see Chapter 2). CruA from *Chl. tepidum* is a lycopene monocyclase.

Alternatively, lycopene could be made asymmetric by isomerization of one of the double bonds away from the central 15, 15′ bond. It has been shown that the products of CrtQ are cis-isomers of lycopene (1, 3, 6, 13), and it is possible that the product of CrtH is not all-trans-lycopene but rather a specific cis-isomer. This is an attractive scenario
because cis-carotenoids are found in PSI of cyanobacteria (17) and have been reported in the reaction centers of GSB (4). In addition, this could explain why CruA from cyanobacteria has not been shown to have lycopene cyclase activity in E. coli (Maresca, Graham, et al., in preparation; 28): the engineered carotenogenic E. coli strains desaturate phytoene using CrtI, whose products are all-trans-carotenoids. The accumulation of all-trans-lycopene in the Chl. tepidum cruA mutant (Chapter 2) can be explained by the fact that cis- to trans-isomerization of carotenoids occurs spontaneously in the light; given that GSB are obligate photoautotrophs, and in the lab are grown at saturating light intensities, this isomerization probably occurs readily.

In contrast to the CruA from Chl. tepidum, CruA from Chl. phaeobacteroides strain 266T does not have any dicyclase activity in E. coli (Figures 3.5C and 3.6C). CruB has dicyclase activity on neurosporene (Figure 3.6D), low dicyclase activity on lycopene (Figure 3.5D), and cyclase activity on γ-carotene (Figure 3.5E). Its high activity on neurosporene could indicate that the branch point between monocyclic and dicyclic carotenoid pathways in isorenieratene-producing species occurs with cyclization of neurosporene by CruA or CruB, a pathway similar to that proposed by Hirabayashi et al. (15) when they observed accumulation of the dicyclic neurosporene derivative 7,8-dihydro-β-carotene in Chl. phaeobacteroides strain 1549. If this were the case, a desaturase capable of recognizing the 7,8-dihydro- bond adjacent to a β ring would be necessary in order to produce β-carotene, β-isorenieratene, and isorenieratene (see Figures 3.10 and 3.11 for structures). Thus, green species would need more than cruB alone to produce isorenieratene, and when cyclization is inhibited, brown species would accumulate neurosporene rather than lycopene.

When cruB is expressed in Chl. tepidum, nearly 100% of the carotenoids produced are dicyclic (Figure 3.7). Although it is possible that one of the genes between BE1 and cruB is a carotenoid 7,8 desaturase, neither of these genes appears to be conserved among all three of the genomes of brown-colored GSB species, and neither has any homology to any known carotenoid biosynthesis gene. Additionally, if the substrate for CruB is neurosporene, the activity in Chl. tepidum would not be as high: neurosporene does not seem to accumulate to detectable levels in wild-type cells, and its monocyclic derivative 7,8-dihydro-γ-carotene is only seen in very low amounts. In fact,
neither this compound nor 7,8-dihydro-β-carotene are observed in the Chl. tepidum strain expressing cruB (Figure 3.7).

When lycopene cyclization is inhibited in the brown-colored species Chl. phaeobacteroides strain 1549 and Pld. phaeum, two different phenotypes are observed. In Pld. phaeum, lycopene accumulates, as would be expected if CruA is a lycopene monocyclase and CruB is a γ-carotene cyclase (Figure 3.8C). In Chl. phaeobacteroides strain 1549, very little lycopene and a large amount of neurosporene accumulate, as would be expected if neurosporene were the substrate for CruA and CruB (Figure 3.8B). However, Chl. phaeobacteroides strain 1549 accumulates primarily neurosporene and neurosporene derivatives during exponential growth, with very little β-isorenieratene and undetectable isorenieratene (Figure 3.8B; 15). It seems more likely not that neurosporene is the substrate for cyclization but that this strain has altered CrtQ activity, such that neurosporene rather than lycopene is produced early in the pathway. If the results for Pld. phaeum are more representative of brown species in general, they support the hypothesis that the branch point between mono- and dicyclic carotenoids occurs when γ-carotene is cyclized by CruB or hydroxylated by CrtC (Figure 3.10). Because neurosporene accumulates both as an end product and an intermediate in Chl. phaeobacteroides strain 1549, the pathway in this species has more branches (Figure 3.11).

Characterization of CruA and CruB has confirmed that CruA is a carotenoid monocyclase found in all GSB, and CruB is a carotenoid cyclase specific to isorenieratene biosynthesis. In addition, these experiments have opened up a new series of questions regarding generation of asymmetric and symmetric carotenoids, the biosynthesis and function of cis-carotenoid isomers in Type I reaction centers, and horizontal versus vertical transmission of biosynthetic pathways among photosynthetic bacteria.
3.5 References


Chapter 3 Figure Legends

Figure 3.1 Structures of chlorobactene, the carotenoid produced by green-colored species of green sulfur bacteria, and isorenieratene, the carotenoid produced by brown-colored species of green sulfur bacteria.

Figure 3.2 Phylogenetic tree of cruA-type lycopene cyclases. Nucleotide sequences were used because the similarity between the protein sequences is so high that resolution of branch order is impossible. The cruA homolog from Synechocystis sp. PCC 6803 is used as an outgroup.

Figure 3.3 Constructs for expression of cruA and cruB from Chl. phaeobacteroides strain 266T in E. coli. (A) cruA in vector pET16b. (B) cruB in vector pCOLA.

Figure 3.4 Plasmid pXN-BEBj-aadA for expression of BE1 and cruB in Chl. tepidum. The gene encoding the BChl c C-20 methyltransferase (CT0028, bchU) is replaced at the start codon of bchU with a gene cluster from Pld. phaeoalcalthratiforme and the aadA cassette conferring resistance to streptomycin and spectinomycin.

Figure 3.5 Elution profiles of pigments produced in E. coli strain BL21(DE3):pACYC that synthesizes lycopene. Absorption spectra on the right correspond to the numbered peaks. (A) BL21(DE3) pACYC; peak 1 is lycopene. (B) BL21(DE3) pACYC + pCTLY encoding cruA from Chl. tepidum. Peak 2 is γ-carotene and peak 3 is β-carotene. (C) BL21(DE3) pACYC + pCPL1 encoding cruA from Chl. phaeobacteroides strain 266T. (D) BL21(DE3) pACYC + pCOLA-B encoding cruB from Chl. phaeobacteroides strain 266T. (E) BL21(DE3) pACYC + pCPL1 + pCOLA-B.

Figure 3.6 Elution profiles of pigments produced in E. coli strain BL21(DE3): pACNEUR that synthesizes neurosporene. Absorption spectra on the right correspond to the numbered peaks. (A) BL21(DE3) pACNEUR; peak 1 is neurosporene. (B) BL21(DE3) pACNEUR + pCTLY encoding cruA from Chl. tepidum. Peak 2 is 7,8-dihydro-β-carotene. (C) BL21(DE3) pACNEUR + pCPL1 encoding cruA from Chl. phaeobacteroides strain 266T. Peak 3 is 7,8-dihydro-γ-carotene. (D) BL21(DE3) pACNEUR + pCOLA-B encoding cruB from Chl. phaeobacteroides strain 266T. (E) BL21(DE3) pACNEUR + pCPL1 + pCOLA-B.

Figure 3.7 (A) Elution profile of carotenoids from wild-type Chl. tepidum. (B) Elution profile of carotenoids from stationary phase Chl. phaeobacteroides strain 1549. (C) Elution profile of carotenoids from Chl. tepidum strain CT0028::BEBj-aadA expressing cruB. Peaks: 1. Chlorobactene. 2. Isorenieratene. 3. β-isorenieratene. 4. β-carotene.

Figure 3.8 Elution profiles of carotenoids in GSB left untreated (black lines) or treated with 20 µM MPTA (gray lines). (A) Wild-type Chl. tepidum. (B) Chl. phaeobacteroides strain 1549. (C) Pld. phaeum. (D) Mixture of E. coli strains producing lycopene and neurosporene. Peaks: 1: OH-chlorobactene glucoside laurate. 2: Chlorobactene. 3. Cis-chlorobactene. 4. γ-carotene. 5. 7,8- dihydro-γ-carotene. 6. Lycopene. 7. β-isorenieratene.

**Figure 3.9** Carotenoid composition of *E. coli* expressing *cruA* from *Chl. tepidum*, left untreated or treated with MPTA. Values are the average from triplicate samples.

**Figure 3.10** Proposed biosynthetic pathway for isorenieratene in *Chl. phaeobacteroides* strain 266\(^T\), *Pld. phaeoclathratiforme*, and *Pld. phaeum*.

**Figure 3.11** Proposed biosynthetic pathway for carotenoids in *Chl. phaeobacteroides* strain 1549. Carotenoids marked with (*) accumulate as end products in this strain.
Table 3.1 Primers used in this work.

<table>
<thead>
<tr>
<th>Organism</th>
<th>Gene/ORF</th>
<th>Purpose</th>
<th>Primer name</th>
<th>Primer sequence</th>
<th>Restriction site</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Pld. phaeoclathratiforme</em></td>
<td><em>ppha2727</em>- <em>ppha2725</em></td>
<td>Expression of <em>ppha2727</em> in <em>C. tepidum</em></td>
<td>ppBE-F10</td>
<td>CAA GTA ATA TTT CCA TGG TAT GAA TAC AAA GAG</td>
<td><em>Nco</em>I</td>
</tr>
<tr>
<td><em>Pld. phaeoclathratiforme</em></td>
<td><em>ppha2727</em>- <em>ppha2725</em></td>
<td>Expression of <em>ppha2727</em> in <em>C. tepidum</em></td>
<td>ppBE-R10</td>
<td>CGA TCA GCC TGG TCA TGT TCT AGA TGG AAC GG</td>
<td><em>Xba</em>I</td>
</tr>
<tr>
<td><em>Pld. phaeoclathratiforme</em></td>
<td><em>ppha2727</em>- <em>ppha2724</em></td>
<td>Expression of <em>ppha2727</em> in <em>C. tepidum</em></td>
<td>ppCRUB-F10</td>
<td>CC GTT CCA TCT AGA ACA TGA CCA GGC TGA TCG</td>
<td><em>Xba</em>I</td>
</tr>
<tr>
<td><em>Pld. phaeoclathratiforme</em></td>
<td><em>ppha2725</em>- <em>ppha2724</em></td>
<td>Expression of <em>ppha2727</em> in <em>C. tepidum</em></td>
<td>ppCRUB-R10</td>
<td>GCG ATG TAA TAG TCG ACT TGT TTG GTA CGG</td>
<td><em>Sal</em>I</td>
</tr>
<tr>
<td><em>Chl. phaeobacteroides</em></td>
<td>cruA/ cpha0241</td>
<td>Expression in <em>E. coli</em></td>
<td>CPL1-F3</td>
<td>AAT CAA TGT TTG TCA TAT GTC ATC ACT GCT</td>
<td><em>Nde</em>I</td>
</tr>
<tr>
<td><em>Chl. phaeobacteroides</em></td>
<td>cruA/ cpha0241</td>
<td>Expression in <em>E. coli</em></td>
<td>CPL1-R1</td>
<td>ATG ACG AAG GAT CCG AAG AGT</td>
<td><em>Bam</em>HI</td>
</tr>
<tr>
<td><em>Chl. phaeobacteroides</em></td>
<td>cruB/ cpha2404</td>
<td>Expression in <em>E. coli</em></td>
<td>CPL2-FN</td>
<td>C GTG ATT TGC CAT GGT AGC</td>
<td><em>Nco</em>I</td>
</tr>
<tr>
<td><em>Chl. phaeobacteroides</em></td>
<td>cruB/ cpha2404</td>
<td>Expression in <em>E. coli</em></td>
<td>CP2407-KR</td>
<td>GCC ATG TCG GTA CCT GTT TAT G</td>
<td><em>Kpn</em>I</td>
</tr>
</tbody>
</table>
Table 3.2 Plasmid combinations in *E. coli* strain BL21(DE3) pACLYC and carotenoids produced. Carotenoids in parentheses are minor products.

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>pUC19 (empty vector)</th>
<th>pCTLY (cruA, Chl. tepidum)</th>
<th>pET16b (empty vector)</th>
<th>pCPL1 (cruA, Chl. phaeobacteroides)</th>
</tr>
</thead>
<tbody>
<tr>
<td>pCOLA-DUET1 (empty vector)</td>
<td>lycopene</td>
<td>γ-carotene (β-carotene)</td>
<td>lycopene</td>
<td>γ-carotene</td>
</tr>
<tr>
<td>pCOLA-B (cruB, Chl. phaeobacteroides)</td>
<td>lycopene</td>
<td>β-carotene</td>
<td>lycopene</td>
<td>β-carotene</td>
</tr>
</tbody>
</table>
Table 3.3 Plasmid combinations in *E. coli* strain BL21(DE3) pACNEUR and carotenoids produced.

<table>
<thead>
<tr>
<th></th>
<th>pUC19 (empty vector)</th>
<th>pCTLY (cruA, Chl. tepidum)</th>
<th>pET16b (empty vector)</th>
<th>pCPL1 (cruA, Chl. phaeobacteroides)</th>
</tr>
</thead>
<tbody>
<tr>
<td>pCOLA-DUET1 (empty vector)</td>
<td>neurosporene</td>
<td>7,8-dihydro-β-carotene</td>
<td>neurosporene</td>
<td>7,8-dihydro-γ-carotene</td>
</tr>
<tr>
<td>pCOLA-B (cruB, Chl. phaeobacteroides)</td>
<td>neurosporene</td>
<td>7,8-dihydro-β-carotene</td>
<td>neurosporene</td>
<td>7,8-dihydro-β-carotene</td>
</tr>
</tbody>
</table>
Figure 3.1

chlorobactene

isorenieratene
Figure 3.2.
Figure 3.3.
Figure 3.4
Figure 3.5
Figure 3.6.
Figure 3.8.
Figure 3.9
Figure 3.10

\[ \text{ζ-carotene} \]

\[ \text{neurosporene} \]

\[ \text{lycopene} \]

\[ \text{γ-carotene} \]

\[ \text{β-carotene} \]

\[ \text{β-isorenieratene} \]

\[ \text{isorenieratene} \]
Figure 3.11.
CHAPTER 4

Identification of two genes encoding carotenoid-modifying enzymes in *Chlorobium tepidum*

Abstract

Chlorobium (Chl.) tepidum produces chlorobactene as its primary carotenoid. Small amounts of chlorobactene are hydroxylated by the enzyme CrtC, then glucosylated and acylated to produce chlorobactene glucoside laurate. The genes responsible for the modification of chlorobactene, CT1987 and CT0967, have been identified by comparative genomics, and the genes were insertionally inactivated in Chl. tepidum to verify their predicted function. The gene encoding chlorobactene glucosyltransferase (CT1987) has been named cruC, and that encoding chlorobactene lauroyltransferase (CT0967) has been named cruD. Homologs of these genes are found in the genomes of all sequenced green sulfur bacteria as well as in the genomes of several non-photosynthetic bacteria that produce similarly modified carotenoids. The other bacteria in which these genes are found are not closely related to green sulfur bacteria or to one another. This suggests that the ability to synthesize modified carotenoids has been a frequently transferred trait, presumably because the production of these carotenoids is physiologically beneficial.
4.1 Introduction

Carotenoids with covalently linked sugars and acylated sugars have been found in many bacteria and occur in a wide range of phyla in both photosynthetic and non-photosynthetic species. Typically, glucose is linked to a hydroxyl group on the carotenoid backbone, and an acyl group may then be esterified to the glucose. The carotenoid-linked sugar may also be mannose (3) or fucose (25, 32, 33), and the length of the acyl chain can vary from 1 to 20 carbons (8). Glycosylated and acylated carotenoids are often found in the cytoplasmic membrane (10, 18, 20, 30), but their physiological functions remain uncharacterized.

Modified carotenoids are synthesized by such a variety of organisms that it is impossible to infer function from distribution. They are found in both aerobic (32) and anaerobic (31, 34, 35, 36) photosynthetic bacteria as well as in the flowers and fruits of a few plant species (8). Among non-photosynthetic bacteria, glycosylated and acylated carotenoids have been identified in aerobes (20) and facultative aerobes (23) as well as in halophiles (21), psychrophiles (10), and thermophiles (38, 39). The thermophilic filamentous anoxygenic phototroph (FAP) *Chloroflexus aurantiacus* produces OH-γ-carotene glucoside esters, for which the major fatty acids are hexadecanoate and hexadecenoate (35, 36). *Thermus thermophilus* has also been shown to produce thermozeaxanthins, which are zeaxanthin glucoside esters (39). *Deinococcus* sp., which are phylogenetically related to *Thermus* sp., are red-pigmented as a result of carotenoid production, and at least two species of glycosylated xanthophyll carotenoids are produced by these organisms (4). Finally, acylated carotenoids also occur in *Myxococcus xanthus*, where they are synthesized in response to blue light (7), nutrient status (11), or heavy metal concentration (26).

In some of these species, it has been proposed that the modified carotenoids affect membrane rigidity in response to stress (10, 37). Experiments with polar carotenoids in artificial membranes suggest that they reduce the rate of oxygen diffusion through the membrane, thus protecting membrane lipids from oxidative damage (29).

The enzymatic pathway for modification of carotenoids has been elucidated in few species. In *Erwinia (Er.) herbicola* and related bacteria, the glucosyltransferase is
CrtX, which is a membrane-associated protein that transfers glucose from UDP-glucose to zeaxanthin (18, 24). *Staphylococcus aureus* has an unrelated carotenoid glycosyltransferase, which converts the C-30 carotenoid 4, 4’-diaponeurosporenic acid to glycosyl 4, 4’-diaponeurosporenoate (27). The next step, acylation, produces staphyloxanthin, or β-D-glucopyranosyl 1-O-(4,4’-diaponeurosporenoate)-6-O-(12-methyltetradecanoate). The genes encoding the glycosyltransferase and acyltransferase in *S. aureus* have been named *crtQ* and *crtO*, respectively (27). Here, the *S. aureus* enzymes will be referred to by their locus tags, SA2350 and SA2352, in order to avoid confusion with ζ-carotene desaturase, which is also named CrtQ (2), and β-carotene ketolase, also named CrtO (17).

The green sulfur bacterium *Chlorobium (Chl.) tepidum* produces a small amount of OH-chlorobactene glucoside laurate (30, 36), which is found in the cytoplasmic membrane rather than associated with the chlorosomes (30). The genes in the chlorobactene biosynthetic pathway up to OH-chlorobactene have previously been identified (Figure 4.1; 12; Chapter 2), but no homologs of the *E. herbicola* glycosyltransferase CrtX could be found in the predicted ORFs in the *Chl tepidum* genome (9). In this study, a comparative genomics approach was used to identify possible carotenoid-specific glycosyltransferases and acyltransferases in *Chl. tepidum* and other carotenogenic organisms. This approach yielded one candidate glycosyltransferase and one candidate acyltransferase, each of which were insertionally inactivated in *Chl. tepidum* to confirm the roles of their products in carotenogenesis. The identification of these genes, which have now been named *cruC* and *cruD*, completes the carotenoid biosynthetic pathway in *Chl. tepidum* (Figure 4.1).

### 4.2 Materials and Methods

#### 4.2.1 Identification of candidate genes and phylogenetic comparisons.

All open reading frames (ORFs) in the *Chl. tepidum* genome that had been annotated as either glycosyltransferases or acyltransferases were compared to the proteins in the non-redundant databases in the Comprehensive Microbial Resource (CMR, The Institute for Genomic Research, Rockville, MD; http://cmr.tigr.org/tigr-scripts/CMR/CmrHomePage.cgi) and
the Integrated Microbial Genomes site (IMG, Joint Genome Institute, Walnut Grove, CA; http://img.jgi.doe.gov/cgi-bin/pub/main.cgi) by using the BLASTP algorithm (1). The genome neighborhoods of the best hits were visually inspected for the presence of other ORFs predicted to be involved in carotenoid biosynthesis. This procedure identified one glycosyltransferase and one acyltransferase in *Chl. tepidum* that were apparently orthologous to genes, which were found in close proximity to genes encoding predicted enzymes of carotenogenesis in other organisms. These two genes, *CT1987* and *CT0967*, were subjected to further analysis. For construction of phylogenetic trees, amino acid sequences were aligned using the ClustalW tool in MacVector 7.2.3 (Accelrys, San Diego, CA) and neighbor-joining trees were generated from the alignments with Paup 4.0 (Sinauer Associates, Inc., Sunderland, MA).

### 4.2.2 Construction of mutant strains.

Constructs for inactivation of *CT1987* and *CT0967* were made by PCR amplification of the 5’ ends of the genes using primer pairs CT1987 FA/RA and CT0967 FA/FB and amplification of the 3’ ends using primer pairs CT1987 FB/RB and CT0967 FB/RB (Table 4.1). PCR products were digested with the indicated restriction enzymes and the *aadA* cassette conferring resistance to streptomycin and spectinomycin was digested with the same enzymes. The resulting fragments were gel-purified using the Eppendorf kit and ligated. The ligation product with the size corresponding to A-*aadA*-B was then gel-purified and used to transform *Chl. tepidum* as previously described (12).

Growth of *Chl. tepidum* wild-type and mutant strains for pigment analysis and growth rate measurements, and HPLC separation and identification of carotenoid species are as described in Chapter 2.

### 4.3 Results

#### 4.3.1 Identification of candidate genes for the terminal steps of carotenogenesis in *Chl. tepidum*.

The *Chl. tepidum* genome encodes nearly 40 glycosyl transferases and several acyltransferases. Rather than systematically attempt to inactivate each gene by interposon mutagenesis to identify the genes whose products are involved in carotenogenesis, we employed a comparative bioinformatics approach to identify a subset of can-
candidate genes. As shown in Figure 4.2, the genome contexts of CT1987 and CT0967 provide no clues that the proteins encoded by these genes might be involved in carotenogenesis, since each is flanked by hypothetical and conserved hypothetical proteins. However, apparently orthologous genes in Deinococcus (De.) geothermalis, Rhodopirellula (Rhp.) baltica, and Frankia sp. EAN1pec are clustered with one or more genes encoding proteins annotated as carotenoid dehydrogenases or phytoene synthase (Figure 4.2). Based upon this evidence, CT1987 and CT0967 were selected for insertional inactivation and further characterization.

4.3.2 Insertional inactivation of CT1987 and CT0976. A DNA fragment encoding the aadA gene, which encodes aminoglycoside acetyltransferase and confers resistance to spectinomycin and streptomycin, was ligated to the DNA fragments encoding the upstream and downstream flanks of CT1987 and CT0967 (Figure 4.3A). The resulting constructions were purified and used to transform wild-type Chl. tepidum cells. Transformants were selected and streaked to obtain isolated colonies and to allow segregation of mutant alleles. PCR across these two genes showed that each transformant had a 1.0-kb insertion in the inactivated gene, which corresponds to the size of the aadA cassette from plasmid pSRA81 (Figure 4.3B). As no product corresponding to the size of the wild-type product was observed in either mutant, we conclude that both mutations were completely segregated.

4.3.3 Characterization of mutant strains. When carotenoids of the wild type strain were analyzed by HPLC, six major carotenoid species were resolved (Figure 4.4). Based upon previous characterization of various crt mutants as well as mass spectrometric analyses (13), these peaks correspond to OH-chlorobactene glucoside (1), OH-chlorobactene (2), OH-chlorobactene glucoside laurate (3), chlorobactene (4), γ-carotene (5), and 1',2', dihydrochlorobactene (6). As shown in Figure 4.4, peaks 1, 2 and 3 were all missing when the chlorobactene hydroxylating enzyme CrtC is missing. The CT1987::aadA mutant no longer produced OH-chlorobactene glucoside or OH-chlorobactene glucoside laurate, but still produced a carotenoid with an absorption spectrum identical to that of chlorobactene and the mass of OH-chlorobactene (550.4 Da). The CT0967::aadA mutant could no longer synthesize OH-chlorobactene glucoside laurate (peak 3) but could still produce OH-chlorobactene glucoside (peak 1) and OH-
chlorobactene (peak 2). The carotenoid analyses of the two mutants clearly show that \textit{CT1987} encodes the OH-chlorobactene glucosyltransferase while \textit{CT0967} encodes the OH-chlorobactene glucoside lauryltransferase. These two ORFs have accordingly been renamed \textit{cruC} and \textit{cruD}.

**4.3.4 Growth rates of mutants.** Previous studies had shown that a \textit{Chl. tepidum} mutant lacking all carotenoids is viable and that mutants producing predominantly lycopene, γ-carotene, and chlorobactene (see Fig. 4.1) can grow at rates close to the wild type (13). The \textit{crtC} mutant, which cannot hydroxylate chlorobactene and therefore makes no modified carotenoids (Figure 4.4), grows at \sim90\% of the rate of wild-type at 70 \si{\mu mol\ photons\ m^{-2}\ s^{-1}} and \sim60\% of the rate of wild-type at low light intensity 10 \si{\mu mol\ photons\ m^{-2}\ s^{-1}} (13). The \textit{cruC} (glucosyltransferase) mutant, which makes OH-chlorobactene, has a similar pattern, whereas the \textit{cruD} (acyltransferase) mutant, which makes OH-chlorobactene glucoside, grows more slowly under all conditions tested (Figure 4.5).

**4.3.5 Sequence comparisons and phylogenetic analyses.** The glycosyltransferases found in regions with other putative carotenoid biosynthesis genes were aligned with each other. Among the carotenoid-specific glycosyltransferases, those from GSB grouped together, with the enzymes from \textit{Cfx. aurantiacus} and \textit{Roseiflexus} sp. strain RS-1 clustering together closely to the GSB. The recently characterized diaponeurosporene glycosyltransferase from \textit{Staphylococcus aureus}, SA 2350 (27), groups with the other putative carotenoid glycosyltransferases, but outside the group of C_{40} carotenoid glycosyltransferases (Figure 4.6). The acyltransferases found in the same or similar regions were aligned together with selected acyltransferases from lipid and amino acid biosynthetic pathways. The probable carotenoid acyltransferases form a phylogenetic cluster separate from the other two (Figure 4.7).

**4.4 Discussion**

This work demonstrates the utility of comparing gene neighborhoods of orthologous genes to identify which may be involved in a particular biosynthetic pathway. In this case, this approach identified single candidates for the carotenoid glucosyltransferase and carotenoid acyltransferase in the genome of \textit{Chl. tepidum}. Inactivation of these two genes and biochemical analyses of the corresponding mutants confirmed that the product of \textit{CT1987} is the OH-chlorobactene glucosyltransferase and that \textit{CT0967} encodes the OH-
chlorobactene glucoside lauroyltransferase. The identification of these genes, now named cruC and cruD, respectively, completes the identification of the genes responsible for the enzymes of the chlorobactene biosynthetic pathway in Chl. tepidum (see Figure 4.1).

For Chl. tepidum, having a population of carotenoids anchored in the cytoplasmic membrane most likely protects the membrane either from radicals generated by the reaction centers, from reactive oxygen species, from chlorophyll triplet states, or from all of these. Although green sulfur bacteria (GSB) are not oxygenic, the strong reductants generated by their photosystem can react readily with O₂ to form reactive oxygen species (14), which have the potential to damage membrane lipids. Given their absolute requirement for light, many GSB inhabit the upper parts of anoxic zones, where periodic disturbance of the chemocline by wind or tides could result in exposure to molecular oxygen. Because homologs of both CruC and CruD are found in the genomes of all sequenced GSB, it is likely that these specialized carotenoids play important role(s) in the cell. Additionally, although neither gene is essential, loss of either gene reduces the growth rate of the cells even under optimum laboratory conditions. This suggests that their role is not limited to protection from stress conditions such as oxygen exposure.

Glycosylated and acylated carotenoids are widespread among both photosynthetic and non-photosynthetic bacteria, so it is unlikely that they are principally involved in light harvesting and the conversion of light energy to chemical energy. Because of their amphipathic nature, they are most likely anchored in the cytoplasmic membrane with a specific orientation, where they serve some structural or protective function such as absorbing and dissipating the energy from reactive oxygen species, ultraviolet light or chlorophyll triplet states. Polar modifications of carotenoids, such as the introduction of hydroxyl or keto groups, tend to interact with the phosphate headgroups of lipids, thereby orienting the carotenoids within membranes (6, 29). The acyl group should anchor these molecules within the membrane or provide structural interactions with certain membrane proteins. These molecules resemble unusual, staple-shaped glycolipids, and specific lipids are known to play important roles in the biogenesis of some membrane proteins (5) or to have specific binding sites on membrane proteins associated with photosynthesis (15, 19). In light of the observation that chlorobactene glucoside esters are found in the cytoplasmic membrane rather than the chlorosome and co-purify with reaction centers (30), it
is possible that these modified carotenoids play a minor role either in biogenesis of the photosystem proteins or in photoprotection of the reaction centers in the membrane.

The five sequenced filamentous anoxygenic phototroph (FAP) species, *Chloroflexus* (Cfx.) *aurantiacus*, *Cfx. aggregans*, *Herpetosiphon aurantiacus*, *Roseiflexus castenholzii* and *Roseiflexus* sp. strain RS-1, also have homologs of both *cruC* and *cruD*. As shown in Figure 4.6, the FAP glycosyltransferases cluster together within the group of carotenoid glycosyltransferases, but their lineage is deeply divergent and not very similar to the GSB lineage. In the case of CruD, the *Cfx. aurantiacus* homolog forms a deeply branching lineage within the putative group of carotenoid acyltransferases (Figure 4.7). These results are similar to those for other *crt* genes (13), and taken together, these data suggest that the carotenoid biosynthesis pathway in FAP species may have been acquired separately from the bacteriochlorophyll (BChl) c and chlorosome biosynthetic pathways, most likely from a different source.

The *S. aureus* carotenoid glycosyltransferase SA2350 aligns with the others, but appears to form an outgroup within the carotenoid glycosyltransferases. The large sequence differences among these proteins may be due to the fact that the substrates for SA2350 are C_{30} carotenoids, rather than the C_{40} compounds synthesized by the other species. Because the *S. aureus* acyltransferase, SA2352, seems to be unrelated to the other putative carotenoid acyltransferases, it is also possible that the similarity of the SA2350 carotenoid glycosyltransferase to the other carotenoid glycosyltransferases arose by convergent evolution from a non-orthologous glycosyltransferase subfamily.

The putative carotenoid acyltransferases form a coherent phylogenetic group when aligned and compared with lipid O-specific or N-specific acyltransferases. This grouping, in conjunction with the gene neighborhood specific analysis, allows us to predict that the ORFs RB11942 (in *Rhp. baltica*) and Der0090 (in *De. radiodurans*), as well as ORF 9 in the *M. xanthus* carotenoid biosynthesis operon, also encode carotenoid acyltransferases. Since the *S. aureus* acyltransferase SA2352 does not align with these carotenoid acyltransferases, CruD appears to be the first characterized member of a novel carotenoid acyltransferase family.

The identification of *cruC* and *cruD*, and the distribution of these genes in other bacteria, highlights the chimeric nature of the chlorobactene biosynthetic pathway in
green sulfur bacteria. The pathway from phytoene to chlorobactene or isorenieratene is more similar to the synechocyanthol pathway in cyanobacteria than to carotenoid biosynthesis in most other bacteria or even isorenieratene biosynthesis in actinomycetes (13, 22). However, the enzymes involved in modification of the cyclized carotenoids (ψ-end hydroxylation, glycosylation, and acylation) are more similar to those in FAPs and non-photosynthetic bacteria (13). This raises interesting questions about the acquisition and evolution of carotenoid biosynthesis genes in GSB. It seems plausible that the carotenoid biosynthesis pathway was acquired in two or more events from different sources. Considering the prevalence of lateral gene transfer in the evolutionary history of carotenoid biosynthesis (16, 28), this is perhaps the rule rather than the exception.
4.5 References


Chapter 4 Figure Legends

Figure 4.1 Biosynthetic pathway of OH-chlorobactene glucoside laurate from chlorobactene. All genes in this pathway have been inactivated in *Chl. tepidum* and the pigments of the mutants have been characterized.

Figure 4.2 Genome neighborhood comparisons. *CT0967* (*cruD*) in *Chl. tepidum* and orthologs in other organisms are gray and white striped; *CT1987* (*cruC*) in *Chl. tepidum* and homologs in other organisms are solid black. Other genes predicted to be involved in carotenoid biosynthesis are light gray; all other genes are white. In two species, orthologs of both *CT1987* and *CT0967* appear in a genomic region with at least one other carotenoid biosynthesis gene. This arrangement of genes suggests that these two ORFs might encode proteins that function in carotenoid biosynthesis.

Figure 4.3. Maps of constructions used to inactivate *CT1987* and *CT0967* and confirmation of segregated mutants. **A.** Strategy and restriction maps of constructions used to inactivate *CT1987* and *CT0967*. **B.** Electrophoretic analysis of PCR products to evaluate putative mutants. The *CT1987* and *CT0967* loci were amplified by PCR using the primers CT 1987 F1 and R1 or CT0967 F1 and R1. The DNA templates used for lanes 2 and 4 were derived from wild type *Chl. tepidum*, for lane 3 from a *CT1987::aadA* transformant, and for lane 5 from a *CT0967::aadA* transformant. Lanes 1 and 6 are DNA size markers (GeneChoice Ladder I).

Figure 4.4. HPLC elution profiles of wild-type *Chl. tepidum* and mutants unable to synthesize various modified carotenoids. Peak 1, OH-chlorobactene (550.4 Da); Peak 2, OH-chlorobactene glucoside (729.4 Da); Peak 3, OH-chlorobactene glucoside laurate (894.6 Da); Peak 4, chlorobactene (532.4 Da); Peak 5, γ-carotene (536.4 Da); Peak 6, 1’,2’-dihydrochlorobactene (534.5 Da). Elution of carotenoid species was monitored at 491 nm.

Figure 4.5. Growth rates of wild type and *crtC*, *cruC*, and *cruD* mutant strains of *Chl. tepidum* at different light intensities

Figure 4.6. Relationships among carotenoid glycosyltransferases. Amino acid sequences were aligned and used to reconstruct the phylogeny; all labels are the automatically assigned locus tags for that ORF. Only *Chl. tepidum* CruC and *S. aureus* SA2350 have been genetically characterized. Another Class II glycosyl transferase from *Chl. tepidum*, CT0362, was used as the outgroup for this comparison

Figure 4.7. Relationships among various acyltransferases. Amino acid sequences were aligned and used to infer phylogenetic relationships. Abbreviations are as follows: *Eco*, *Escherichia coli*; *RB*, *Rhodopirellula baltica*; *Mxx*, *Myxococcus xanthus*; *CP*, *Chl. phaeobacteroides* 266T; *Ctep*, *Chl. tepidum*; *Caur*, *Cfx. aurantiacus*; *Der*, *Deinococcus*
radiodurans; Hin, Haemophilus influenzae; ECA, Erwinia carotovora subsp. atroseptica; HD, Haemophilus ducreyi; Sty, Salmonella typhimurium. Bold lines indicate proteins whose functions have been genetically or biochemically confirmed.
Table 4.1 Primers used in this work

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Primer sequence</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>CT1987 F1</td>
<td>CAG GAA GCT TAT GAA AAG TGC</td>
<td>Confirming segregation</td>
</tr>
<tr>
<td>CT1987 R1</td>
<td>CGC GAG CTA CCG GAG GGT TGG</td>
<td>Confirming segregation</td>
</tr>
<tr>
<td>CT1987 F2</td>
<td>TGG AGT CGA CGA AAG CAT CC</td>
<td>Amplifying right flank of inactivation construct</td>
</tr>
<tr>
<td>CT1987 R2</td>
<td>CTG CCT TAG GTA CCG CGG TTT CG</td>
<td>Amplifying left flank of inactivation construct</td>
</tr>
<tr>
<td>CT1987 F3</td>
<td>CCA GAT TTT TCA CCA CAG GCT CC</td>
<td>Amplifying left flank of inactivation construct</td>
</tr>
<tr>
<td>CT1987 R3</td>
<td>TTC CAC TTT TCC GTC CAG AGC C</td>
<td>Amplifying right flank of inactivation construct</td>
</tr>
<tr>
<td>CT0967 F2</td>
<td>GAA TGC ACT GTG GAC TTT TCC G</td>
<td>Amplifying right flank of inactivation construct</td>
</tr>
<tr>
<td>CT0967 R2</td>
<td>AGT GAA GAA CTG GTA CCT GAA GAG</td>
<td>Amplifying left flank of inactivation construct</td>
</tr>
<tr>
<td>CT0967 F3</td>
<td>TCA TCA ATG CGT GCG ACA ACT</td>
<td>Amplifying left flank of inactivation construct</td>
</tr>
<tr>
<td>CT0967 R3</td>
<td>CAT CGT TTC TTC TAC GGA TTT ACC C</td>
<td>Amplifying right flank of inactivation construct</td>
</tr>
</tbody>
</table>

*KpnI sites are underlined, and SalI sites are italicized.*
Figure 4.1

chlorobactene $\rightarrow$ CruC

OH-chlorobactene $\rightarrow$ CruC

OH-chlorobactene glucoside $\rightarrow$ CruD

OH-chlorobactene glucoside laurate
Figure 4.2

- *Chl. tepidum*
- *De. geothermalis*
- *Rhp. baltica*
- *Frankia sp. EAN1pec.*
- *Chl. tepidum*
Figure 4.3
Figure 4.6

S. aureus SA2350
Rhp. baltica RB11941
  Rfx. castenholzii Rcas0436
  Rfx. strain RS-1 RS1_2830
  Cfx. aggregans Cagg0670
  Cfx. aurantiacus Caur1259
   Hps. aurantiacus Haur3309
  De. geothermali s Dgeo0680
   Chl. phaeobacteroides Cpha2756
   Chl. ferrooxidans cfer0710
   Chl. tepidum CruC
   Chl. chlorochromatii Cag0212
    Frankia sp. EAN1pec fran1539
    Chl. tepidum CT0362

0.1 changes
CHAPTER 5

Identification of the bacteriochlorophyll c C-20 methyltransferase in *Chlorobium tepidum*

ABSTRACT

Bacteriochlorophylls (BChls) c and d, the major antenna bacteriochlorophylls in green-colored species of green sulfur bacteria, differ only by the presence of a methyl group at the C-20 methine bridge position in BChl c. A gene potentially encoding the C-20 methyltransferase, bchU, was identified by comparative analysis of the Chlorobium (Chl.) tepidum and Chloroflexus aurantiacus genome sequences. Homologs of this gene were amplified and sequenced from Chl. phaeobacteroides strain 1549, Chl. vibrioforme strain 8327d, and Chl. vibrioforme strain 8327c, which produce BChl e, d, and c respectively. A one-nucleotide insertion in the bchU gene of Chl. vibrioforme strain 8327d was found to cause a premature stop codon and thus the formation of a truncated protein. A spontaneous mutant of this strain that produces BChl c (strain 8327c) has a second frame-shift mutation that restores the correct reading frame in bchU. The bchU gene has been inactivated in Chl. tepidum, a BChl c-producing species, and the resulting mutant produced only BChl d. Strains producing BChl c grow faster than those with BChl d in low-light conditions. The bchU gene encodes the C-20 methyltransferase in BChl c and e biosynthesis in Chlorobium species, and methylation at the C-20 position to produce BChl c rather than BChl d confers a significant competitive advantage to green sulfur bacteria living at limiting light intensities.
5.1 Introduction

The light-harvesting antenna system of green sulfur bacteria, the chlorosome, is uniquely adapted to collecting light in low-light environments (4, 5). With up to 215,000 molecules of bacteriochlorophyll (BChl) per chlorosome and up to 250 chlorosomes per cell (23), some green sulfur bacteria (GSB) are able to harvest light energy and grow at light fluxes as low as 3 nmol photons m\(^{-2}\) s\(^{-1}\) (27). GSB produce BChl c, d, or e as primary antenna pigments, in addition to small amounts of BChl a, which is associated with the CsmA chlorosome baseplate protein, the FMO antenna protein, and the reaction center (21). BChls c, d, and e differ in their methylation patterns: BChl e has a formyl group rather than a methyl group at the C-7 position, and a methyl group at the C-20 position (16). BChls c and d both have methyl groups at the C-7 position, but only BChl c has a methyl group at the C-20 position (18, 29; see Figure 5.1).

The deep water or shallow sediments in which GSB are found are usually light-limited, with typical light intensities ranging from 0.1 to 10 µmol photons m\(^{-2}\) s\(^{-1}\) (15, 39, 40). In nature, brown-colored BChl e-containing organisms are frequently found at lower light intensities than those containing BChl c or d (19, 24, 28, 42). Studies of a BChl d-producing strain of Chlorobium (Chl.) vibrioforme, which spontaneously began to synthesize BChl c when grown at low light intensity, suggest that BChl c may be a more effective antenna molecule than BChl d at low light intensity (6, 8, 32). These studies showed that, in Chl. vibrioforme strain 8327d, the addition of the C-20 methyl group to BChl d shifts the absorption maximum of the Q\(y\) peak to the red from 732 to 751 nm in whole-cell spectra and from 658 to 667 nm in methanol extracts (8; see table 5.1).

The complete genome sequence of Chl. tepidum, a BChl c-producing green sulfur bacterium, has provided the basis for investigating the genes involved in pigment and chlorosome biosynthesis (10, 12). Because a bchK mutant that completely lacks BChl c and yet is able to grow photoautotrophically has been constructed, genes encoding any enzymes in the BChl c biosynthetic pathway can in principle be inactivated (14). In addition, because BChl d, which lacks the C-20 methyl group, is a common pigment in GSB, loss of the gene encoding the C-20 methyltransferase should not be a lethal mutation. In
fact, having isogenic strains that differed only in their pigment composition allowed physiological experiments that clarified why one pigment may be preferred over the other \textit{in vivo}. Here, we demonstrated that the product of the gene $CT0028$, now named $bchU$, is responsible for methylation of the C-20 methine bridge in BChl $c$ and probably for that of BChl $e$ from brown-colored \textit{Chlorobium} species. We also demonstrate that this C-20 methylation confers a competitive advantage to GSB when they are growing at low light intensity.
5.2 Materials and Methods

5.2.1 Strains and growth conditions. Wild-type Chl. tepidum and Chl. phaeobacteroides strain 1549 were grown as described in Chapter 2. Chl. vibrioforme strains 8327d and 8327c (8) were grown in CL medium as for Chl. tepidum, at a temperature of 30°C.

5.2.2 Inactivation of CT0028. A 1.1-kb fragment of Chl. tepidum open reading frame CT0028 was amplified from genomic DNA by using forward primer CT0028-FE and reverse primer CT0028-BP, which contain an EcoRI recognition site and a PstI recognition site, respectively (Table 2). Plasmid pCFT was obtained by digesting this PCR product with EcoRI and PstI and cloning it into the multiple cloning site of pUC19 digested with the same enzymes. Plasmid pCFT-X was made by insertion of the aadA cassette from pSRA2 (13) between internal KpnI sites within the coding sequence of the CT0028 gene.

5.2.3 Chlorosome preparation and analysis. Cultures (2 L) of Chl. tepidum wild type and the bchU mutant were grown at ~150 µmol photons m⁻² s⁻¹. Chlorosomes were isolated and purified from these cultures according to the procedure described previously (9, 38) and stored at -80°C until used. For the comparison of molar extinction coefficients, chlorosomes were thawed on ice in the dark. Wild-type chlorosomes were diluted 1:10000 in cell-harvesting buffer (10 mM KH₂PO₄, 50 mM NaCl, pH 7.0) or 1:100 in acetone:methanol (7:2, v/v), then 1:100 in 100% methanol. Chlorosomes from the bchU mutant were diluted 1:5000 in cell harvesting buffer or 1:100 in acetone:methanol (7:2, v/v) and then 1:50 in 100% methanol. The absorption spectra of the resulting solutions were scanned immediately from 350 to 900 nm. The concentrations of BChl c or d were determined from the absorbance maxima of the methanol extracts, using the following extinction coefficients: BChl c: ε₆₆₇ = 86.0 L g⁻¹ cm⁻¹; BChl d: ε₆₅₆ = 82.3 L g⁻¹ cm⁻¹ (31). Extinction coefficients were converted to units of L mol⁻¹ cm⁻¹ assuming that 33% of the BChl is [8-propyl, 12-ethyl] BChl and 67% is [8-ethyl, 12-ethyl] BChl, giving a molar extinction coefficient for BChl c of ε₆₆₇ = 69,717 L mol⁻¹ cm⁻¹ and for BChl d of ε₆₅₆ = 65,974 L mol⁻¹ cm⁻¹. The calculated concentration was then used to calculate the molar extinction coefficients of the BChls in the chlorosomes.
5.2.4 Growth rate measurements. Chl. tepidum was grown at 46°C and 146 or 8 μmol photons m$^{-2}$ s$^{-1}$. and Chl. vibrioforme was grown at 31°C and 70 or 2 μmol photons m$^{-2}$ s$^{-1}$. Cultures were grown on a rotating wheel that was uniformly illuminated from the front. Before measuring growth rates, cultures were grown to late exponential phase at the light intensity at which the growth rate was to be measured. Cultures were then diluted to an OD$_{600\text{ nm}}$ of 0.1, and the optical density was monitored until the culture reached stationary phase. Because of the transient appearance of polysulfide and elemental sulfur granules, which affects the light scattering in the tubes, measurements below an OD$_{600\text{ nm}}$ ~0.8 were not used when determining the growth rate.

5.2.5 Competition experiments. The BChl $c$- and $d$-containing strains of Chl. tepidum or Chl. vibrioforme were mixed in CL medium with no antibiotics to produce cultures in which the BChl $c$-containing strain represented 10%, 20%, 50%, 80%, and 90% of the total cells. The mixed cultures were grown under the same conditions described above. Cultures grown at the lower light intensity were incubated for 36 hours, diluted 1:20, and incubated again for 18 hours; this incubation time and light intensity corresponds to approximately 12 generations. Cultures grown at the higher light intensity were incubated for 24 h, diluted 1:20, and incubated for another 24 h; this incubation time and light intensity corresponds to a total of approximately 20 generations for the wild type. The Chl. vibrioforme cultures grown at higher light were incubated for 26 h (~7.8 generations for both strains); those at lower light for 126 h, or 6 generations for the BChl $d$ strain and 8 for the BChl $c$ strain.

Because the amount of BChl in each strain is similar, the ratio of the two strains in the culture at the end of the experiment can be determined by measuring the ratio of BChl $c$ to BChl $d$. Aliquots (1.0 ml) were removed from each culture before the experiment began and at the end of the experiment. Pigments were extracted from the cells by sonication in 0.4 ml acetone:methanol (7:2, v/v), and the absorption spectrum of the extracts was measured from 350 to 900 nm. The ratio of BChl $c$ to BChl $d$ in each sample was calculated using the equations $A_{656} = \varepsilon_{c,656} \times c_c + \varepsilon_{d,656} \times c_d$ and $A_{667} = \varepsilon_{c,667} \times c_c + \varepsilon_{d,667} \times c_d$ where $A$ is the value of the absorbance, $\varepsilon$ is the extinction coefficient and $c$ is the concentration. Extinction coefficients for the absorption maxima of the pigments in methanol.
were the same as those above; $\varepsilon_{c,656}$ and $\varepsilon_{d,667}$ were calculated to be 52.4 and 21.9 L g$^{-1}$ cm$^{-1}$, respectively.
5.3 Results

5.3.1 Identification of genes potentially encoding the BChl c C-20 methyltransferase.

The *Chl. tepidum* genome contains many methyltransferase genes, and since most of the genes involved in BChl biosynthesis are not organized in large gene clusters or as operons, the identification of the BChl c C-20 methyltransferase was not straightforward. In the biosynthesis of cobalamin, CbiL catalyzes the methylation of the C-20 methine bridge of cobalt-precorrin 2 (31, 33). *Chl. tepidum* appears to encode all of the genes necessary to synthesize cobalamin (10), and this includes a gene (*CT0388*) encoding a protein with strong sequence similarity to the CbiL proteins of other bacteria. In addition to this gene, the genome encodes a gene, *CT1763*, for a protein that is weakly similar to CbiL, which was considered to be a possible candidate for the C-20 methyltransferase in BChl c biosynthesis. The *CT1763* gene, which encodes a protein with similarity to both CbiL and uroporphyrin-III C-methyltransferase, was insertionally inactivated and a fully segregated mutant strain was obtained. However, this mutant still produced BChl c and had no apparent pigmentation differences or other phenotypic differences from the wild type strain (N.-U. Frigaard and D.A. Bryant, personal communication).

Genes whose products are involved in the same biosynthetic pathway, whose products are found in the same cellular compartment, or whose products are coregulated are often clustered together in operons on bacterial genomes. As noted above, although such clustering is not pronounced in the *Chl. tepidum* genome, analyses of the chlorophyll biosynthesis genes in the green filamentous bacterium *Chloroflexus (Cfx.) aurantiacus* showed more evidence for clustering (12). In *Cfx. aurantiacus*, BChl biosynthesis and chlorosome production are induced when cells are shifted from oxic to anoxic conditions in the light (11, 26). In this organism, genes encoding several chlorosome envelope proteins are clustered with *bchK*, the BChl c synthase, and a gene encoding a methyltransferase that had been annotated as *crtF* (25, 46; Figure 5.2A). The *crtF* gene encodes an O-methyltransferase in the sphaeroidenone biosynthetic pathway of purple bacteria (1). The genome of *Chl. tepidum* contains a gene (*CT0028*) which encodes a methyltransferase with very strong sequence similarity to the product of the *Cfx. aurantiacus* “CrtF” protein. However, since neither *Cfx. aurantiacus* nor *Chl. tepidum* synthesizes O-
methylated carotenoids (36, 37, 38), it was clear that both genes annotated as \textit{crtF} had been mis-identified. The presence of this methyltransferase gene adjacent to \textit{bchK}, known to be involved in BChl \textit{c} biosynthesis, and clustered with genes encoding structural components of the chlorosome, strongly suggested that this gene could encode a methyltransferase involved in BChl \textit{c} biosynthesis.

Broch-Due and Ormerod (8) found that prolonged growth of \textit{Chl. vibrioforme} strain 8327\textit{d} at low light intensity led to gain-of-function mutants, which had acquired the ability to methylate BChl \textit{d} and thus produce BChl \textit{c}. Similar results were also reported by Smith and coworkers (6, 20), and the experiments of Broch-Due and Ormerod were repeated by Tamiaki and coworkers (32). These results show that the wild-type strain of \textit{Chl. vibrioforme}, strain 8327\textit{d}, carries the gene for the BChl \textit{c} C-20 methyltransferase, but that this gene is inactive. The apparently low frequency at which the gain-of-function mutants appeared in the various studies of three groups over a period of 25 years suggests that the gene encoding the methyltransferase was mutated to an inactive form but that the function of this gene could be restored by a secondary mutation. By sequencing the methyltransferase genes from several strains, which included \textit{Chl. vibrioforme} strain 8327\textit{d} and 8327\textit{c}, this prediction was directly tested.

\textbf{5.3.2 Sequence analysis of \textit{bchU} and \textit{crtF} genes.} Degenerate primers were designed based on regions conserved between \textit{crtF} in \textit{Rhodobacter capsulatus}, \textit{bchU} in \textit{Cfx. aurantiacus}, and \textit{CT0028} in \textit{Chl. tepidum} (Table 5.2). These primers were used to amplify and sequence \textit{CT0028} homologs from \textit{Chl. tepidum}, \textit{Chl. vibrioforme} strains 8327\textit{d} and 8327\textit{c}, and \textit{Chl. phaeobacteroides} strain 1549. Using the translated sequences of these homologs, as well as the sequences of previously published \textit{crtF} genes, a phylogenetic tree was made comparing the amino acid sequences of \textit{CT0028} homologs with \textit{CrtF} homologs. The \textit{CT0028} sequences from green sulfur bacteria and \textit{Cfx. aurantiacus} cluster in a group that is clearly distinct from the \textit{CrtF} sequences of purple bacteria (Figure 5.2B).

The nucleotide sequences of these genes showed that \textit{Chl. vibrioforme} 8327\textit{d} has a single-nucleotide insertion 260 nucleotides downstream from the translation start site of the gene. This frame-shift insertion results in an in-frame stop codon 78 nucleotides
downstream of the mutation (Fig. 5.3), the result of which would be a peptide of irrelevant sequence that is only 87 amino acids long, or less than one-fourth of the length of the 338-amino-acid protein of the wild-type. The BChl c-producing strain of \textit{Chl. vibrioforme} (strain 8327c) has a deletion of an adenine residue in the same region, returning the gene to the correct reading frame. These results provide a molecular explanation for the observed shift in BChl type in \textit{Chl. vibrioforme} strain 8327d cells cultured under low-light selection conditions.

5.3.3 Construction and verification of a \textit{CT0028} mutant of \textit{Chl. tepidum}. Wild-type \textit{Chl. tepidum} was transformed with linearized plasmid pCFT-X, which contained the \textit{CT0028} gene interrupted with the \textit{aadA} antibiotic resistance marker (Fig. 5.4A). Green transformant colonies appeared on selective plates containing spectinomycin and streptomycin within 5 days. Segregation of the mutant and wild-type alleles was confirmed by PCR using the primers CT0028F1 and CT0028B1 to amplify the \textit{bchU} locus from genomic DNA extracted from the wild-type and mutant strains (Fig. 5.4B). The product amplified from wild-type \textit{Chl. tepidum} cells was 1.0 kb, while the product amplified from the \textit{CT0028} mutant cells was 1.9 kb; this result confirmed that the \textit{CT0028} gene had been insertionally inactivated with the \textit{aadA} cassette. No 1.0-kb fragment corresponding to the wild-type \textit{CT0028} allele was amplified from the mutant strain. Together, these results indicate that the segregation of the \textit{CT0028} and \textit{CT0028::aadA} alleles was complete and that the \textit{CT0028} gene is not essential under the growth conditions employed.

5.3.4 Pigment analysis of \textit{Chl. tepidum CT0028 (bchU)} mutant. Pigments extracted with acetone-methanol (7:2 v/v) from both the wild type and \textit{CT0028} mutant strain of \textit{Chl. tepidum} were separated on a C-18 reverse-phase HPLC column. The elution profiles of the two strains demonstrated a similar distribution of BChl homologs, indicating that methylation at the C-8 and C-12 positions is not affected by the \textit{CT0028} mutation (Fig. 5.5A). All BChl c homologs in the wild type had identical absorption spectra with a maximum at 667 nm, whereas the major BChl species in the \textit{CT0028} mutant had absorption spectra identical to BChl d with a maximum at 654 nm. The BChl d homologs eluted approximately 1.5 minutes earlier than their BChl c counterparts; this indicates that the BChl c homologs are slightly more hydrophobic, as expected. Mass spectrometry of the
BChl pigments showed that each of the four major BChl homologs produced by the mutant are lighter than the corresponding Bchl c compound in the wild-type by 14 mass units. This change in mass indicates the loss of a single methyl group from each BChl homolog in the CT0028 mutant (Fig. 5.5). These results confirm that the product of CT0028 is the BChl c C-20 methyltransferase; accordingly, this gene has been renamed bchU.

The BChl c- or BChl d-to-protein ratio was determined for cells of the wild type and bchU mutant strain of Chl. tepidum that had been grown to an OD_{600 nm} of ~1.0 at high (707 µmol photons m^{-2} s^{-1}) and low (10 µmol photons m^{-2} s^{-1}) light intensity, respectively. As previously reported for Chl. tepidum (7), the cellular content of antenna BChl increased as the growth light intensity decreased. The antenna BChl:protein ratio (w/w) of the two strains was identical (0.058 ± 0.007) in cells grown at high light intensity. However, at low light intensity, the antenna BChl:protein ratio of the wild type (0.174±0.013) was 21% greater than that of the bchU mutant strain (0.143±0.007). The BChl c:BChl a and BChl d:BChl a ratios were also determined by HPLC analysis of pigments extracted from the same cells. These measurements showed that the ratio of BChl c:BChl a (54.2:1) in the wild-type cells grown at high light intensity was almost identical to the ratio of BChl d:BChl a (53.6:1) in the bchU mutant. However, at low light intensity, the ratio of BChl c:BChl a in the wild-type cells was 57:1 and the ratio of BChl d:BChl a in the bchU mutant cells was 41:1, a 28% decrease.

5.3.5 Analysis of molar extinction coefficients of BChl c and d. The absorption spectrum of BChl c in methanol has maxima at 667 and 435 nm, while that of BChl d has maxima at 654 and 428 nm (34). When the spectra of the pigments in organic solvents are adjusted to reflect their molar absorptivities (Fig. 5.6A), it is clear that BChl c has greater absorbance in both the blue and red regions of the spectrum. Integration of the absorption spectra from 350 to 800 nm shows that BChl c has about 15% greater overall absorbance in this wavelength range than BChl d. When the absorption spectra in aqueous buffer of chlorosomes isolated from the wild type and the bchU mutant are plotted to reflect their respective molar absorptivities (Fig. 5.6B), several differences can be observed for the chlorosomes containing these two BChl species. In the case of Chl.
tepidum, the bandwidth of the Qy peak of chlorosomes containing BChl c is 59 nm, whereas that of chlorosomes containing BChl d is 43 nm. The Soret band of chlorosomes containing BChl c is 6% higher than that of chlorosomes containing BChl d, although this band is not detectably broader. In addition, BChl c absorbs slightly more than BChl d in the region between 550 and 650 nm. Integration of the spectra in the wavelength range from 650 nm to 850 nm shows that chlorosomes containing BChl c have 25% greater absorption than chlorosomes containing BChl d. When one integrates these spectra in the wavelength range from 350 to 800 nm (Fig. 5.6B), these spectral differences result in ~16% greater overall absorbance by BChl c-containing chlorosomes than BChl d-containing chlorosomes.

5.3.6 Growth characteristics of BChl c and BChl d strains. Growth rate measurements of both BChl c- and d-producing strains of Chl. tepidum and Chl. vibrioforme were performed at nearly optimal growth temperatures and several light intensities. The growth rates and doubling times determined from these experiments are summarized in Table 5.3. At saturating light intensity, the BChl c and BChl d strains of both organisms grew at nearly identical rates (Table 5.2, Fig. 5.7, Fig. 5.8). At limiting light intensities, the BChl c-producing strain of Chl. vibrioforme grew 30% faster than the BChl d-producing strain, while in Chl. tepidum the wild-type strain grew 50% faster than the bchU mutant. These results are similar to those previously reported for Chl. vibrioforme (8). The BChl d-producing strain of Chl. vibrioforme was shown to grow faster than the BChl c isolate at very high light 8); however, when Chl. tepidum was grown at 580 µmol photons m^{-2} s^{-1}, an inhibitory light intensity for the wild-type strain, the BChl d-producing strain still grew more slowly than wild type (Fig. 5.7A).

5.3.7 Competition between BChl c- and d- containing strains. On the basis of the determined growth rates of the two strains, the BChl c strain would be expected to grow to much higher density than the BChl d strain within 12 generations if a mixed culture were grown at low light intensity, but not if it were grown at saturating light intensity. To test this prediction, the wild type and the bchU mutant strain of Chl. tepidum were mixed in selected ratios and grown at 146 or 8 µmol photons m^{-2} s^{-1}. As expected, the ratio of BChl c to BChl d in mixed cultures grown at high light intensity remained essentially un-
changed over 20 generations (Fig. 5.9B), indicating that the two strains were equally able to harvest and transfer light energy under those conditions. Under the conditions of these experiments, the growth rate difference between the two strains must have been less than 0.5%. However, in the mixed cultures grown at 8 µmol photons m\(^{-2}\) s\(^{-1}\), the proportion of BChl c in the culture after 12 generations was much higher than it had been in the inoculum at the onset of the growth experiment (Fig. 5.9A). In those cultures in which 80 or 90% of the initial inoculum had been the BChl c- producing strain, >99% of the pigment in the culture was BChl c after 12 generations at low light intensity. In the cultures in which the BChl c strain constituted between 10 % and 50% of the initial inoculum, between 67 % and 95% of the pigment in the culture was BChl c after 12 generations (Fig. 5.9A). Similar results were obtained with mixed cultures of Chl. vibrioforme str. 8327c and 8327d grown at 70 µmol photons m\(^{-2}\) s\(^{-1}\) (Figure 5.10). The results obtained in the Chl. tepidum competition growth experiments could be accurately predicted solely on the basis of the growth rates of the individual strains measured at the light intensities employed (Figure 5.9). This indicates that the results are solely due to the growth rate differences that result from the differences in the BChl contents in the chlorosomes of the two strains.
5.4 Discussion

The results presented here establish that the CT0028 gene, now denoted bchU, of Chl. tepidum encodes the BChl c C-20 methyltransferase. This gene was amplified and sequenced from three species (Chl. tepidum, Chl. vibrioforme, and Chl. phaeobacteroides) that methylate the C-20 position of their major BChl. In a BChl d-producing strain of Chl. vibrioforme, the gene is present but inactive because of a single nucleotide insertion which changes the reading frame of the gene and gives rise to a premature, in-frame stop codon that results in a truncated and inactive polypeptide. A secondary frameshift mutation has reactivated this gene in Chl. vibrioforme 8327c, which is a gain-of-function mutant derived from the “wild-type” parental strain. Finally, when the same gene was insertionally inactivated in Chl. tepidum, which produces BChl c, the resulting mutant strain only produced BChl d.

As reported previously, the C-20 methyl group is derived from S-adenosyl methionine (20). In the cobalamin biosynthetic pathway of Salmonella typhimurium, the C-20 methyltransferase, CbiL, only methylates substrates that have cobalt inserted into the corrin ring, and may require this metal for catalysis of the reaction (33). Although BchU and CbiL are not very similar in sequence, they may have a similar catalytic mechanism, as proposed on the basis of the recently obtained crystal structures of BchU and CbiL from Chl. tepidum (44, 45). The reaction is proposed to follow an S\(_{N2}\)-like mechanism in which a tyrosine conserved in both CbiL and BchU is first deprotonated; the tyrosinate anion is proposed to act as the base that deprotonates the C-19 position to generate a carbanion at the C-20 position. This carbanion then attacks the methyl group of S-adenosylmethionine, which leads to the transfer of the methyl group to the C-20 position of BChl c (45).

The distribution of BChl d homologs in the bchU mutant of Chl. tepidum is virtually identical to the distribution of BChl c homologs in the wild type (Fig. 5.5), which suggests that the absence of methylation at the C-20 position does not affect the activities of the C-8 and C-12 methyltransferases in the bchU mutant (BchQ and BchR, respec-
tively; 12, 13). Additionally, in *Chl. tepidum* strains in which the C-8 and C-12 methyltransferases have been inactivated (A. Gomez Maqueo Chew, N.-U. Frigaard, and D. A. Bryant, manuscript in preparation), all BChl c homologs are methylated at the C-20 position. These observations suggest either that the C-20 methylation occurs before the other methylations or that BchU can methylate a wide range of substrates. Additionally, recent work by Harada *et al.* suggests that BchU cannot methylate protoporphyrin IX, but can methylate bacteriochlorophyllide *d* regardless of the functional group at the C-3 position (17). In the purple photosynthetic bacterium *Rhodobacter capsulatus*, CrtF has broad substrate specificity (2); the ability to methylate a broad range of substrates might be characteristic of the CrtF and BchU families of methyltransferases.

Only one green sulfur bacterium, *Chl. limicola* UdG6040, has been reported to accumulate both BChl *c* and *d* (3, 35). In this organism, the ratio of BChl *d* to BChl *c* increases when the culture is growing very rapidly; this observation suggests that the C-20 methyltransferase in this organism may not be very efficient and that its activity may be rate-limiting. This observation also suggests that C-20 methylation occurs before the esterifying long-chain alcohol “tail” is added and that BchU is not able to methylate BChl *d*. However, if BChl *d* is rapidly imported into chlorosomes, it might not be available to BchU, since this molecule would be sequestered within the chlorosome envelope.

*Chl. tepidum* and *Chl. vibrioforme* strains unable to methylate the C-20 position of their antenna BChls grew significantly more slowly at low light intensities. The *bchU* mutant of *Chl. tepidum* grew as fast as the wild-type (>99% of the rate of wild-type) at saturating light intensities, but had a growth rate that was only 66% of that of the wild-type when the light intensity was reduced to 8 μmol photons m<sup>-2</sup> s<sup>-1</sup>. When BChl *c* - and *d*-producing strains of *Chl. tepidum* were grown in direct competition, the ratio of BChl *c* to BChl *d* did not change when the cultures were grown at high light intensity; this indicates that the two strains had identical growth rates under these conditions. However, at low light intensity (8 μmol photons m<sup>-2</sup> s<sup>-1</sup>), BChl *c* predominated after only 12 generations of growth even in cultures that initially had a BChl *c*:BChl *d* ratio of 1:10. Very similar results were obtained using BChl *c* - and *d*-producing strains of *Chl. vibrioforme* (Figures 5.9 and 5.10).
These results explain why, as has been previously reported, *Chl. vibrioforme* is able to revert to producing BChl *c*. The cultures from which revertants were isolated had been grown at low light intensity for periods of several months to several years (8, 20, 32), thus applying significant selective pressure. Any organism that deleted a single nucleotide in the vicinity of the original insertion could produce BChl *c* and outcompete the “wild-type” BChl *d*-producing strain under the culture conditions employed. Interestingly, the BChl *d* strain of *Chl. vibrioforme* strain 8327 was reported to grow faster at high light intensity than the revertant BChl *c* strain (8), but the same was not true of the *Chl. tepidum* strains grown at a photoinhibitory light intensity (580 µmol photons m⁻² s⁻¹). The *bchU* mutant grew at ~92% of the rate of wild type under these high light intensity, photoinhibitory conditions (Figure 5.7A).

Wild-type *Chl. tepidum* grows faster at low light intensity than the *bchU* mutant for two reasons. Firstly, BChl *c* has enhanced absorption relative to BChl *d* throughout the physiologically relevant wavelength range of 350 to 850 nm (see Fig. 5.6). The *Q₅* absorption peak of chlorosomes containing BChl *c* is significantly broader than that of chlorosomes containing BChl *d*; moreover, the Soret absorption and the absorption in the remainder of the visible region of BChl *c* are also greater than for BChl *d* (see Fig. 5.6B). Both in organic solvents and in isolated chlorosomes, these small differences add up to a 16% difference in total absorption. However, the greatest differences occur in the absorbance region from 650 to 850 nm that includes the *Q₅* absorption band (Fig. 5.6B). Since the growth experiments were performed with incandescent illumination that has relatively little intensity in the blue region of the spectrum, it can easily be seen that the large difference in *Q₅* absorption would provide a strong selective advantage. Secondly, although the cellular content of antenna BChl was identical for the wild-type and *bchU* mutant strains of *Chl. tepidum* grown at high light intensity, the cellular content of BChl *d* in the *bchU* mutant was significantly smaller (21%) than the content of BChl *c* in the wild type when cells were grown at low light intensity. In combination, the enhanced absorption per BChl molecule and greater number of antenna molecules per cell permit the wild type to grow ~50% faster than the *bchU* mutant at low light intensity.
The fact that isogenic BChl d and c strains have nearly identical growth rates at intermediate and high light intensities leads to the question of why, if BChl c is a more effective light-harvesting molecule over a wider range of light intensities, does BChl d exist at all? One possible explanation for this may be related to the observation that Chl. tepidum does not grow well at very high light intensities (>300 µmol photons m\(^{-2}\) s\(^{-1}\)) (14); this observation suggests that at high light intensity chlorosomes are unable to deliver harvested light energy efficiently to the reaction centers. Under such conditions, a less efficient light-harvesting antenna might be more useful. If this were the case, BChl d-containing strains would grow faster than BChl c-containing strains at high light intensity; this was observed in Chl. vibrioforme (8), but has not been seen in Chl. tepidum.

Field studies have consistently shown that the majority of green-colored green sulfur bacterial species found in nature produce BChl d (40); however, the majority of cultured species produce BChl c. This difference may reflect the culture conditions employed during initial enrichments: if green sulfur bacteria are enriched for and isolated under low-light intensity conditions with an incandescent light source, it will be the case that BChl c-producing species will probably be strongly selected.

Field studies in aquatic systems have also consistently shown a pattern of vertical stratification of green sulfur bacterial species (26, 19, 24, 42), with BChl d-containing species closer to the top of the water column, BChl e-containing brown colored species closer to the bottom, and BChl c-containing species in between. The predominance of BChl e-containing species at greatest depth is probably due to the enhanced absorption of BChl e in the blue-green wavelength region of the visible spectrum, the wavelength range of light that penetrates most deeply in aquatic environments (42, 43). In a study of 41 lakes, Vila and Abella (42) found that green-colored (BChl c- or BChl d-containing) green sulfur bacteria predominated when red and near-infrared wavelengths were represented in the light reaching the metalimnion. As noted above, BChl c absorbs approximately 16% more than BChl d in the wavelength range from 350 to 850 nm; however, in the red and near-infrared region, this difference increases to 25%. Thus, BChl c would be even more advantageous in such natural environments. Vila and Abella (42) additionally reported that brown-colored, BChl e-containing green sulfur bacteria predominated in
environments in which green-yellow light was available and longer wavelengths were absent. Since the proportion of red and near-infrared light decreases with increasing depth, the brown-colored green sulfur bacteria will usually be found at greatest depth in the water column.

The stratification pattern in sediments is slightly different. Brown-colored (BChl e-producing) species are only rarely found in sediments (27), where the downwelling irradiation is relatively enriched in red wavelengths (22, 29). In sandy sediments, infrared (IR) and near-IR radiation is attenuated less quickly than shorter wavelengths (22, 29); in this environment, BChl c-producing green-colored species might be expected to predominate; however, both green-colored and brown-colored Prosthecochloris species have been isolated from shallow sediments (Martinez-Alonso et al., 2005; Nicholson et al., 1987).

Although the availability of light is an important factor in determining which types of green sulfur bacteria will develop in a given environment, it is not the only important factor. The stratification of differently pigmented species can also reflect the geochemistry of the system in which these organisms live. The lower region of the euphotic zone in freshwater lakes often has a significantly lower redox potential than its upper part, and this low-light, high-sulfide region is usually dominated by the BChl e-containing green sulfur bacteria. Since the green species are more commonly found in low-light environments with slightly higher redox potentials (43), it may sometimes be that in a region in which the light conditions alone would favor either BChl e- or c-producing strains, the chemistry of the water favors the brown-colored, BChl e-containing strains, which out-compete the green bacteria and thus predominate.
5.5 References


Table 5.1. Absorbance properties of major bacteriochlorophyll species in Chl. tepidum and Chl. vibrioforme

<table>
<thead>
<tr>
<th>Strain</th>
<th>Soret-maximum (acetone)</th>
<th>Q&lt;sub&gt;y&lt;/sub&gt; maximum (acetone)</th>
<th>Soret-maximum (in vivo)&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Q&lt;sub&gt;y&lt;/sub&gt; maximum (in vivo)&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chl. tepidum (BChl c)</td>
<td>435</td>
<td>667</td>
<td>460</td>
<td>751</td>
</tr>
<tr>
<td>Chl. tepidum AbchU (BChl d)</td>
<td>428</td>
<td>655</td>
<td>451</td>
<td>736</td>
</tr>
<tr>
<td>Chl. vibrioforme 8327c (BChl c)</td>
<td>435</td>
<td>667</td>
<td>462</td>
<td>750</td>
</tr>
<tr>
<td>Chl. vibrioforme 8327d (BChl d)</td>
<td>428</td>
<td>656</td>
<td>446</td>
<td>733</td>
</tr>
</tbody>
</table>

<sup>a</sup> *In vivo* refers to the value for intact cells.
Table 2. Sequences of oligonucleotides used for cloning and sequencing of \( bchU \).

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Sequence (5’ – 3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CT0028 FD</td>
<td>CAC AGG GCC AAC GAG CTG GTC TT(T/C) AA(A/G)</td>
</tr>
<tr>
<td>CT0028 BD</td>
<td>CAC AGC CTG AAC GAG CAG GTG (A/G)TC (A/G)TA</td>
</tr>
<tr>
<td>CT0028 FE(^a)</td>
<td>CGG GGT GAA TTC GGA CAG GCT GGA TAA C</td>
</tr>
<tr>
<td>CT0028 BP(^b)</td>
<td>GCC GAG CAC CCT GCA GGG CAT TCC</td>
</tr>
<tr>
<td>CT0028 F1</td>
<td>TGA GCA ACA ATGACC TCC TGA ACT A</td>
</tr>
<tr>
<td>CT0028 B1</td>
<td>ACA GCC TGA ACG AGC AGG TG</td>
</tr>
</tbody>
</table>

\(^a\) Underlined sequence is \( \text{EcoRI} \) recognition site for cloning.

\(^b\) Underlined sequence is \( \text{PstI} \) recognition site for cloning.
Table 5.2. Growth rates of *Chl. tepidum* and *Chl. vibrioforme* under varying light conditions.

<table>
<thead>
<tr>
<th>Light intensity</th>
<th>Strain</th>
<th>8 <em>a</em></th>
<th>146 <em>a</em></th>
<th>580 <em>a</em></th>
<th>2 <em>a</em></th>
<th>70 <em>a</em></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>WT</td>
<td><em>bchU</em></td>
<td>WT</td>
<td><em>bchU</em></td>
<td></td>
</tr>
<tr>
<td>Growth rate (h⁻¹)</td>
<td></td>
<td>0.095</td>
<td>0.063</td>
<td>0.295</td>
<td>0.283</td>
<td>0.239</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.221</td>
</tr>
<tr>
<td>Doubling time (h)</td>
<td></td>
<td>6.9</td>
<td>10.7</td>
<td>2.3</td>
<td>2.5</td>
<td>2.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.1</td>
</tr>
<tr>
<td>Fraction of WT</td>
<td></td>
<td>1</td>
<td>0.66</td>
<td>1</td>
<td>0.96</td>
<td>1</td>
</tr>
<tr>
<td>growth rate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.92</td>
</tr>
</tbody>
</table>

*a* Unit: μmol photons m⁻² s⁻¹

*b* The growth rates reported are the averages of two separate experiments in which each rate was determined in triplicate. The standard error was less than 4% in all experiments.
Figure 5.1. Structures of bacteriochlorophylls (BChls) d (A), c (B), and e (C). Carbons are numbered in (A) and are the same for all BChls. Functional group R₁ may be ethyl, methyl, isobutyl, or neopentyl; R₂ may be ethyl or methyl.
**Figure 5.2.** A. Photosynthesis-related gene cluster in *Cfx. aurantiacus*. The gene encoding *bchU*, a *crtF* homolog, is immediately upstream of the BChl *c*-specific synthase and in a region that also contains three genes encoding chlorosome envelope proteins. B. Phylogenetic tree of CrtF (carotenoid O-methyltransferase) protein sequences and BchU (BChl *c* C-20 methyltransferases) from different species. The quinone C-methyltransferase UbiE was used as an outgroup.
Figure 5.3. Alignment of BchU sequences from different strains. A one-nucleotide insertion (indicated by arrows) causes a frame-shift mutation in the *bchU* gene of *Chl. vibrioforme* strain 8327d, resulting in a premature stop codon and a non-functional gene product.
Figure 5.4. **A.** Construct for inactivation of *bchU* in *Chl. tepidum*. The *aadA* cassette from pSRA2 was inserted between internal KpnI sites in the gene. **B.** Confirmation of segregation of the mutant. Lane M, marker (Ladder I from GeneChoice); Lane 1, *bchU* region PCR-amplified from wild-type *Chl. tepidum* cells; Lane 2, *bchU* region PCR-amplified from *bchU* mutant. The amplicon is 0.9 kb longer because of the insertion of *aadA* into *bchU*. 
Figure 5.5. Elution profiles of BCHls from (A) wild-type and (B) bchU mutant cells, monitored at 667 and 654 nm, respectively. Numbers next to peaks indicate mass of the compound in that peak; each peak in the bchU mutant is 14 mass units lighter than the corresponding peaks in wild-type cells and elutes ~1 minute earlier, indicating loss of one methyl group.
Figure 5.6. A. Molar absorption coefficients of BChls c (solid line) and d (dotted line) in methanol. B. Molar absorption coefficients of *Chl. tepidum* chlorosomes containing BChl c (solid line) or d (dotted line) in aqueous buffer. The molar absorption coefficient of BChl c in chlorosomes is ~17% greater overall than that of BChl d.
Figure 5.7. Growth rates of wild-type *Chl. tepidum* (black circles) and *bchU* mutant (gray triangles) at different light intensities. **A.** 580 µmol photons m⁻² s⁻¹. **B.** 146 µmol photons m⁻² s⁻¹. **C.** 10 µmol photons m⁻² s⁻¹.
**Figure 5.8.** Growth rate measurements of wild-type (black triangles) and BChl c revertant (gray circles) strains of *Chl. vibrioforme* at different light intensities.  

A. 70 µmol photons m$^{-2}$ s$^{-1}$.  

B. 2 µmol photons m$^{-2}$ s$^{-1}$.  

---

162
Figure 5.9. Direct growth competition between BChl c- and d-producing strains of *Chl. tepidum*. The fraction of the total pigment content representing BChl c (solid black line) and the predicted fraction (dashed grey line), were calculated as described in the Materials and Methods. 

A. The change in BChl c composition after 20 generations of growth at 146 μmol photons m$^{-2}$ s$^{-1}$

B. The change in BChl c composition after 12 generations of growth at 8 μmol photons m$^{-2}$ s$^{-1}$. 
Figure 5.10. Direct growth competition between BCHl c- and d-producing strains of Chl. vibrioforme. The fraction of the total pigment content representing BCHl c (solid black line) and the predicted fraction (grey line) after 26 generations of growth at 70 μmol photons m⁻² s⁻¹ were calculated as described in the Materials and Methods.
CHAPTER 6:

OXIDATION OF THE C7\(^1\) POSITION OF BACTERIOCHLOROPHYLL E
Abstract

Brown-colored species of green sulfur bacteria have been shown to live at light intensities as low as 3 nmol photons m$^{-2}$ s$^{-1}$. In these species, the antenna bacteriochlorophyll (BChl) is BChl $e$, which is $7^1$ formyl BChl $c$. The addition of this formyl group red-shifts the Soret peak of the BChl and increases its absorbance relative to that of the $Q_y$ peak. Because only brown-colored species make BChl $e$, and three of the twelve available green sulfur bacterial genomes are of brown species, it was possible to use an \textit{in silico} subtractive hybridization analysis to identify potential genes encoding the enzymes required to produce BChl $e$. These genes, including a putative radical SAM protein, a putative short-chain dehydrogenase, and a third homolog of the BChl C$3^1$ hydratase, were found in a gene cluster that also includes the isorenieratene-specific $\gamma$-carotene cyclase \textit{cruB}. Individual genes in this cluster from \textit{Pelodictyon phaeoclathratiforme} as well as the entire 6.5 kb region were inserted into the chromosome of the genetically amenable species \textit{Chlorobium tepidum} to determine their functions.
6.1 Introduction

Chlorobium (Chl.) phaeobacteroides strain BS-1, the green sulfur bacterial (GSB) species that lives 100 m below the surface of the Black Sea (16, 22) is, like all other known GSB, strictly photoautotrophic. In situ, the light flux is approximately 3 nmol photons m$^{-2}$ s$^{-1}$, enough for this species to grow with an estimated doubling time of almost 3 years in the summer, and approximately 26 years in the winter (22; J. Overmann, personal communication). Chl. phaeobacteroides strain BS-1 is a brown species of GSB, which makes isorenieratene as its primary carotenoid and bacteriochlorophyll (BChl) e as its antenna BChl. Brown species of GSB tend to be found in the deepest part of the photic zone, where the remaining light is enriched in yellow and green wavelengths (30, 31). It is this range of wavelengths that BChl e absorbs more efficiently than any other BChl.

All of the brown-colored species of GSB make BChl e, and all of these also make isorenieratene, a carotenoid with two aromatic rings (see Chapter 3). No green (BChl c- or d- producing) species has been shown to make isorenieratene; these species produce chlorobactene as their primary carotenoid. Thus, conversion of a green-colored species to a brown-colored species or vice versa requires acquisition or loss of multiple genes in two unrelated biosynthetic pathways. This observation is worth noting because all of the major phylogenetic groups of GSB have both green and brown members (1; figure 6.1), suggesting that such coordinated genetic changes have occurred several times.

Biosynthesis of BChl e rather than BChl c requires oxidation of the C7$^1$ position of BChl c to a formyl rather than a methyl group (Figure 6.2). Oxidation of the C7$^1$ position of tetrapyrrole molecules occurs in several different systems. In plants, chlorophyll (Chl) a is oxidized at the C7$^1$ position to Chl b by the action of chlorophyll a oxygenase (CAO; 10, 27). This enzyme is a Rieske non-heme iron oxygenase which hydroxylates the C7$^1$ methyl group twice, then water is eliminated to produce a C7$^1$ formyl group (8, 20). In Bacillus (B.) subtilis, heme O can be oxidized at the C8$^1$ position to produce heme A by CtaB, which is also an O$_2$-dependent reaction (4, 26). However, since GSB are obligate anaerobes, it is unlikely that the enzyme(s) involved in this oxidation reaction are oxygenases.
As photoautotrophs living in energy-limited environments, GSB have very small genomes with very little extraneous DNA (9). Because the central metabolism – sulfide oxidation, carbon fixation via the reverse TCA cycle, and type I reaction center-based photosynthesis (13, 18, 23, 32, 33) – of all of these species is so similar, it is probable that a large percentage of the unique genes found in each species plays roles in the adaptations of those species to their particular niches. Using subgroups which share particular phenotypes and identifying the genes that those subgroups share should produce short lists of ORFs that can be combed for ORFs likely to be involved in adaptation to that niche or phenotype.

The genomes of three brown-colored species of GSB are available: Chlorobium (Chl.) phaeobacteroides strain 266T, Chl. phaeobacteroides strain BS-1, and Pelodictyon (Pld.) phaeoclathratiforme strain BU-1. Chl. phaeobacteroides strain 266T is a planktonic rod-shaped GSB strain and Pld. phaeoclathratiforme forms net-like branching filaments. Because the environmental niches populated by these two species are not similar, and because they have such divergent morphologies, the majority of their shared genes are probably either important in the central metabolism or are related to the phenotypes that they have in common, including pigmentation. These two genomes are the largest of the green sulfur bacterial genomes, which range in size from 1.8 to 3.3 Mbp. The genome of Chl. phaeobacteroides strain BS-1 is the composite genome sequence of two closely related but non-identical strains, and is contaminated with non-GSB DNA. For this reason, the comparative analysis was carried out using only Chl. phaeobacteroides strain 266T and Pld. phaeoclathratiforme.
6.2 Methods

6.2.1 Genomic comparisons. The predicted ORFs and their amino acid sequences were downloaded from the Integrated Microbial Genomes website, part of the Joint Genome Institute of the U.S. Department of Energy (http://img.jgi.doe.gov/cgi-bin/pub/main.cgi). Three custom Perl scripts were written and used in the comparative analysis. First, for each pair of GSB genomes, the best blastp hit (2) for each predicted ORF in one genome was identified in the second genome, then the analysis was repeated using the ORFs in the second genome as the query sequences, and lists of best hits were compiled. Second, the two lists for each pair were compared, and lists of those ORFs that were reciprocal best hits were generated. Third, using the list of reciprocal best hits from the two brown species *Chl. phaeobacteroides* strain DSM 266^T^ and *Pld. phaeoclasratiforme* strain BU-1 as a starting point, any ORF in this list with a reciprocal best hit in any other genome was removed. The resulting list of 95 ORFs was inspected and predicted transposases, DNA modifying enzymes, NTPases, and other proteins obviously not involved in pigment biosynthesis were removed.

6.2.2 Dot-blot analysis of distribution of putative BChl e- and isorenieratene-specific genes in a culture collection. Short digoxygenin (DIG)-labeled fragments were amplified from *Chl. phaeobacteroides* strain DSM 266^T^ using primers cruB-F and cruB-R or BE1-F and BE1-R (Table 6.1) and the PCR DIG probe synthesis kit (Roche, Catalog no. 11-636-090-910). 10 ng DNA extracted from 15 species of green sulfur bacteria from the culture collection of Dr. Jörg Overmann (Ludwig-Maximilians University, Munich, Germany) (Table 6.2) was denatured by boiling and applied by vacuum to a pre-baked nylon membrane. The probe was hybridized to the genomic DNA and detected by chemiluminescence according to the manufacturer’s instructions.

6.2.3 Expression of genes from *Pld. phaeoclasratiforme* in *Chl. tepidum*. Plasmid pCFT-X (chapter 5), encoding *bchU* insertionally inactivated with the *aadA* cassette (Sp/Sm^R^), was converted into a platform for expression of other genes by insertion of an NcoI site at the start codon of *bchU*. The restriction site was generated by PCR amplification of the whole plasmid using primers bchU-Nru and NCO-R (Table 6.1),
polishing the ends of the PCR product with Vent polymerase, ligation of the plasmid to itself, and amplification of the resulting plasmid, pCFTN, in *E. coli*. The *csmH* locus was made into a similar platform by amplification of 0.9kb upstream of the *csmH* start codon and 0.7 kb downstream of *csmH* using primer pairs csmHA-F/csmHA-R and csmHB-F/csmHB-R, respectively (Table 6.1). These fragments were inserted into pUC19 using the indicated restriction sites to generate plasmid pUC-HAB.

The putative BChl *e* and isorenieratene-specific genes were amplified from *Pld. phaeocladratiforme* (Table 6.1) and inserted into plasmid pCFTN (Chapter 5; figure 6.3C) or plasmid pUC-HAB using NcoI and KpnI or XbaI sites, generating plasmids pCFTXN-BEDHj and pUCHAB-BEDHj. The *aadA* cassette from plasmid pSRA2 (11) was then inserted into the KpnI sites of the plasmids and the resulting constructs, pCFTXN-BEDHj-aadA and pHA-BEDHj-aadA-HB, were amplified in *E. coli*. To express *bchF3* in *Chl. tepidum*, the fragment encoding the gene *bchF3* and the region between *bchF3* and *BE1* was amplified using primers PPHA-XF1 and PPHA-FNR (Table 6.1) and digested with *NcoI*. After digestion of pHA-BEDHj-aadA-HB with *NcoI*, the fragment encoding *bchF3* was ligated into the *NcoI* site of the vector. The plasmids from the resulting transformants were screened by restriction digestion to identify those in which *bchF3* and *BE1* were in the same orientation, as they are on the chromosome. The plasmids were linearized with Scal or AhdI as appropriate and the linear DNA was used to transform *Chl. tepidum* as previously described (11).

**6.2.4 Pigment analysis.** For analysis of BChls, the method described in Chapter 5 was used.

**6.3 Results**

**6.3.1 “In silico” subtractive hybridization.** *Chl. phaeobacteroides* strain DSM 266<sup>T</sup> and *Pld. phaeocladratiforme* share 1813 ORFs (figure 6.4A), of which 1718 have reciprocal best hits in green-colored species of GSB. Of the remaining 95 ORFs, 38 are hypothetical proteins with no functionally characterized homologs in any species, 8 are transposases, 12 are DNA modifying enzymes, 4 have homology to NTPases, and 13 are unique in the NCBI non-redundant database (figure 6.4B). Among the conserved hypothetical proteins with homologs in cyanobacteria is the pair *cpha2407/ppha2724*, which encode *cruB*, the
isorenieratene-specific carotenoid cyclase. Among the ORFs with some homology to known proteins, one in particular stood out. This pair, cpha2404/ppha2727 (denoted BE1), encodes a putative radical-SAM protein with homology to biotin synthase, BioB (figure 6.7). Like the C7 hydroxylase, BioB catalyzes oxidation of a methyl group. Additionally, these genes are part of a 6.5 kbp gene cluster conserved in both species that includes a bchF homolog, a putative dehydrogenase, a putative aldolase, and cruB (figure 6.5).

6.3.2 Distribution of BE1 and cruB among brown-colored species of GSB. All brown-colored species of GSB have genes that hybridize to the BE1 and cruB probes. The only exception, Chl. phaeovibrioides strain 2631, is also the only brown-colored strain reported to synthesize chlorobactene rather than isorenieratene (J. Overmann, personal communication). None of the green-colored species tested had any gene that hybridized to either probe (Table 6.2).

6.3.3 Expression of genes from Pls. phaeoclathratiforme in Chl. tepidum. After transfer to selective plates, colonies of transformants resistant to streptomycin and spectinomycin appeared in less than a week. All of the strains in which bchU had been replaced by the genes of interest produced BChl d as their only antenna chlorophyll, demonstrating that the mutation had segregated completely. The HPLC elution profiles (Figure 6.10), in vivo absorption spectra, and absorption spectra of whole-cell acetone:methanol (7:2, v/v; data not shown) of these mutants were indistinguishable from the bchU mutant (see Chapter 5). The strains in which csmH had been replaced produced exclusively BChl c, and the pigments produced by these strains were identical to those produced by wild-type Chl. tepidum (Figures 6.8 and 6.9).
6.4 Discussion

The analysis used to identify genes shared by and unique to *Chl. phaeobacteroides* strain 266^T^ and *Pld. phaeoclathratiforme* strain BU-1 was based on reciprocal BLAST hits. Every pair of ORFs that were a best match regardless of which one was used as the query was identified as a “reciprocal best hit.” The reciprocal best hits between *Chl. phaeobacteroides* strain 266^T^ and *Pld. phaeoclathratiforme* strain BU-1 were used as the starting point and compared with the catalogs of reciprocal best hits between those two species and all of the green-colored species. Any ORF that had a reciprocal best hit in the genome of any green-colored species was removed from the list. This type of analysis has been used to refute the hypothesis that *Planctomycetes* are more related to Archaea and eukaryotes than are most bacteria (14), as well as to analyze the phylogenetic relationships between hyperthermophilic archaea (15) and is a commonly used analysis tool when comparing genomes (24, 28, 29).

After comparison of the genomes of the brown-colored species *Chl. phaeobacteroides* strain 266^T^ and *Pld. phaeoclathratiforme* strain BU-1 with the green-colored species *Chl. limicola* f. *thiosulfatophilum, Chl. chlorochromatii, Chl. vibrioforme, Pld. luteolum, Prosthecochloris aestuarii* and *Chl. tepidum*, 95 ORFs were identified as shared by both brown-colored species and no green-colored species (Figure 6.4A). One of these encodes *cruB*, the isorenieratene-specific carotenoid cyclase (see Chapter 3). The presence of this gene in the list indicated that the analysis successfully identified genes specific to and shared by *Chl. phaeobacteroides* strain 266^T^ and *Pld. phaeoclathratiforme*. In all three of the brown-colored GSB species for which there are now genome sequences, *cruB* lies in a small gene cluster that includes three other genes with the same distribution (Figure 6.5).

Of these four genes that have orthologs only in *Chl. phaeobacteroides* strain 266^T^, *Chl. phaeobacteroides* strain BS-1, and *Pld. phaeoclathratiforme* strain BU-1, *cruB* has been expressed in *Chl. tepidum* and *E. coli* and its activity characterized (see Chapter 3). One is a homolog of the BCHl C3^1^ hydratase, BchF (6). GSB all have BchF and a paralog of BchF, now named BchV, which appears to introduce a hydroxyl group with S-stereochemistry at the C3^1^ position (A. Gomez Maqueo Chew and D.A. Bryant, in
preparation). The brown-colored species have a third homolog of \textit{bchF} (denoted \textit{bchF3}; Figure 6.6) in the region close to \textit{cruB}.

The second gene in this cluster is a homolog of BioB currently denoted \textit{BE1} (Figure 6.7). This gene appeared to be a good candidate because it has a cysteine motif characteristic of enzymes that use S-adenosylmethionine (SAM) to generate a radical (25). In aerobic Chl and BChl biosynthesis, the isocyclic ring is formed by the di-iron oxygenase AcsF; in anaerobes, the same reaction is performed by the SAM-dependent enzyme BchE (21). It is possible that conversion of the C7\textsuperscript{1} methyl to formyl follows a similar pattern. The Chl \textit{a} oxidase (CAO) from plants is, like AcsF, an iron-dependent oxygenase, and the protein that performs the same reaction anaerobically could use a radical mechanism. Additionally, the reaction catalyzed by BioB is insertion of a sulfur into dethiobiotin (19), a reaction rather similar to oxidation of a methyl group on a tetrapyrrole.

The last gene in this cluster that is unique to and shared by all three of the genomes of brown species is homologous to the yeast isoprenylcysteine carboxyl methyltransferase Ste14p. This is a eukaryotic protein involved in post-translational modification of other proteins (3). Homologs of Ste14p have been reported to be metalloenzymes (3, 7), though none has been reported to be oxygen-sensitive. Although its function is not obvious from its sequence, the distribution of this gene and its location in three genomes suggest that it might be worth further investigation.

In this study, two of these three genes were inserted into either the \textit{csmH} locus or the \textit{bchU} locus at the predicted start codons of the gene. The \textit{csmH} and \textit{bchU} genes were chosen as platforms because they are expressed constitutively and their absence does not affect growth rate of the cells (12, 17; see Chapter 5). Additionally, the C7\textsuperscript{1}-formylated form of BChl \textit{d}, BChl \textit{f}, has never been seen in nature, and an engineered strain capable of producing such a pigment would be useful for understanding why this BChl is so rare. However, neither \textit{BE1} nor \textit{bchF3} had any effect on BChl content in \textit{Chl. tepidum} when inserted into the \textit{csmH} platform (Figures 6.8 and 6.9), and when inserted into the \textit{bchU} platform, the cells produced only BChl \textit{d} (Figure 6.10). It is possible that the lack of phenotype is due to problems with the heterologous expression system. These particular genes may be expressed inefficiently from the foreign promoter or not expressed strongly.
enough. It is also possible that the start codons of either the *Chl. tepidum* genes or the *Pld. phaeoclatratiforme* genes were predicted incorrectly, and transcription should start in a different location altogether.

Neither of the constructs produced included the putative isoprenylcysteine carboxyl methyltransferase. It is possible that this protein, which has only low similarity to the well-characterized yeast isoprenylcysteine carboxyl methyltransferase Ste14p, carries out the first step in the conversion of the C7 methyl to C7\(^1\) formyl; in this case, no phenotype would be observed. Thus far, it can be concluded that under the conditions tested, *BE1* alone, *BE1* in conjunction with the dehydrogenase downstream, and *BE1* in conjunction with *bchF3* are all unable to oxidize the C7\(^1\) methyl group to a formyl. In fact, none of these performs any chemical reaction on the BChl, as addition of any functional groups would affect the elution times of the pigments even if it had no effect on the absorption spectra.

If indeed the gene or genes necessary for BChl\(e\) biosynthesis are in this gene cluster, it will constitute a small, potentially mobile element that encodes genes involved in both carotenoid and BChl biosynthesis pathways specific to brown-colored species of GSB. Receipt of this genetic element confers an immediate physiological change, the ability to grow at lower light intensity and/or absorb different wavelengths of light. In most of the environments where GSB are found, this would provide a selective advantage. This gene arrangement might also explain why pigment phenotypes are not monophyletic traits. Transfer of this element could easily confer the ability to synthesize both BChl\(e\) and isorenieratene on the recipient strain.
6.5 References


Chapter 6 Figure Legends

**Figure 6.1** Polyphyletic distribution of brown (dark boxes) and green (light boxes) phenotypes. Every major phylogenetic group except the *Chloroherpeton* group has both green and brown members. Modified from Alexander *et al.*, 2002.

**Figure 6.2** Structures of bacteriochlorophylls *c* (top) and *e* (bottom). BChl *c* has a methyl group at the C7 position and BChl *e* has a formyl group.

**Figure 6.3** Plasmids used for creating *Chl. tepidum* strains capable of expressing putative BChl *e*-specific genes from chromosomal loci. Genes to be expressed replace *csmH* (*CT1417*) or *bchU* (*CT0028*) at their start codons. A. Plasmid pHA-nBE-aadA-HB for expression of putative radical SAM protein “BE1”. B. Plasmid pHA-PXE-aadA-HB for expression of *bchF3* and *BE1*. C. Plasmid pCFTXN-BEDHj-aadA for expression of *BE1* and the downstream dehydrogenase.

**Figure 6.4** Results of reciprocal BLAST hit analysis. A. 95 open reading frames are found in both genomes of brown-colored species and do not have reciprocal best hits in the genome of any green-colored species. B. Predicted functions of the 95 ORFs unique to brown-colored species.

**Figure 6.5** Gene neighborhoods around *cruB* (#3) in *Chl. phaeobacteroides* strain BS-1, *Chl. phaeobacteroides* strain 266^T^, and *Pld. phaeoclathratiforme*. Numbered genes are described in Table 6.2.

**Figure 6.6** Relationships between green sulfur bacterial *bchF* homologs. The *bchF* sequence from the filamentous anoxygenic bacterium *Roseiflexus* sp. RS-1 was used as an outgroup.

**Figure 6.7** Phylogenetic tree of selected proteins in the radical SAM family, including putative BChl *e*-specific proteins related to BioB.

**Figure 6.8** Elution profiles of BChls from wild-type *Chl. tepidum* (A) and three isolates of the *csmH::nBE-aadA* strain (B, C, D). The elution times are identical and the absorption spectra of the peaks (not shown) are all normal BChl *c*.

**Figure 6.9** Elution profiles of BChls from (A) wild-type *Chl. tepidum*, (B) wild-type *Pld. phaeum*, (C, D) cells from different transformation plates of strain *csmH::PXE-aadA*. The elution times and absorption spectra (not shown) of the BChls from the mutant strain are identical to those of the BChls from wild-type.

**Figure 6.10** Elution profiles of BChls from (A) wild-type *Chl. tepidum* and (B, C, D) three isolates of the *bchU::nBEDH-aadA* strain. The elution times and absorption spectra of the BChls from the mutant strains are identical to those from the *bchU* mutant (see Chapter 5).
### Table 6.1 Primers used in this work.

<table>
<thead>
<tr>
<th>Organism</th>
<th>Primer name</th>
<th>Purpose</th>
<th>Primer sequence</th>
<th>Restriction site</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Pld. phaeoclathratiforme</em></td>
<td>ppBE1-FN2</td>
<td>Expression of <em>ppha2727</em> in <em>Chl. tepidum</em></td>
<td>CTA TGC CAT GGA TAC AAA GAG AGT AAT TAC AGA GG</td>
<td>NcoI</td>
</tr>
<tr>
<td></td>
<td>ppBE1-Racc</td>
<td></td>
<td>ACG GGA GGC GTA GAC AAG TTC G</td>
<td>Accl</td>
</tr>
<tr>
<td></td>
<td>ppDH-R10</td>
<td>Expression of <em>ppha2727</em> and <em>ppha2726</em> in <em>Chl. tepidum</em></td>
<td>CGA TCA GCC TGG TCA TGT TCT AGA TGG AAC GG</td>
<td>XbaI</td>
</tr>
<tr>
<td></td>
<td>ppha-XF1</td>
<td>Expression of <em>ppha2728</em> in <em>Chl. tepidum</em></td>
<td>TCC CAA ATG CCA TGG GTT GTT AC</td>
<td>NcoI</td>
</tr>
<tr>
<td></td>
<td>ppha-FNR</td>
<td>Expression of <em>ppha2728</em> in <em>Chl. tepidum</em></td>
<td>CCT CTG TAA TTA CTC TCT TTG TAT CCA TGG CAT AG</td>
<td>NcoI</td>
</tr>
<tr>
<td><em>Chl. phaeobacteroides 266</em></td>
<td>BE1-F</td>
<td>Amplification of DIG-labeled fragment to probe culture collection</td>
<td>TCG AGG TTA CCG ATC CAG AC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>BE1-R</td>
<td>Amplification of DIG-labeled fragment to probe culture collection</td>
<td>TCA ACA CCC TCCC AGC ATA CA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CRUB-F</td>
<td>Amplification of DIG-labeled fragment to probe culture collection</td>
<td>ATA GCA GAT AAA GTT CAA ACG</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CRUB-R</td>
<td>Amplification of DIG-labeled fragment to probe culture collection</td>
<td>GCT TCC CTG AAG GAC AAC C</td>
<td></td>
</tr>
</tbody>
</table>
Table 6.2 Genes possibly specific to BChl $e$ biosynthesis in the region around $cruB$ in the genomes of 3 brown-colored GSB

<table>
<thead>
<tr>
<th>Number in Figure 6.4</th>
<th>ORF numbers</th>
<th>annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$cpha_2403$</td>
<td>2-vinyl bacteriochlorophyllide hydratase ($bchF$)</td>
</tr>
<tr>
<td></td>
<td>$ppha_2728$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$cphamn1_3312$</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>$cpha_2404$</td>
<td>Hypothetical protein (radical SAM motif)</td>
</tr>
<tr>
<td></td>
<td>$ppha_2727$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$cphamn1_3313$</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>$cpha_2407$</td>
<td>Conserved hypothetical protein ($cruB$)</td>
</tr>
<tr>
<td></td>
<td>$ppha_2724$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$cphamn1_3315$</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>$cpha_2465$</td>
<td>Isoprenylcysteine carboxyl methyltransferase</td>
</tr>
<tr>
<td></td>
<td>$ppha_2758$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$cphamn1_3318$</td>
<td></td>
</tr>
</tbody>
</table>
**Table 6.3** Distribution of *BE1* and *cruB* in green and brown species in culture collection of Jorg Overmann.

<table>
<thead>
<tr>
<th>Species name</th>
<th>Color</th>
<th><em>cruB</em> present</th>
<th><em>BE1</em> present</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Chl. limicola</em> strain 1402</td>
<td>Brown</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td><em>Chl. phaeobacteroides</em> strain DSM 266&lt;sup&gt;T&lt;/sup&gt;</td>
<td>Brown</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td><em>Chl. phaeovibrioides</em> strain 2631</td>
<td>Brown</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td><em>Pld. phaeocladratiforme</em> strain BU-1</td>
<td>Brown</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Br.Glub.N</td>
<td>Brown</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>E2P3</td>
<td>Brown</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Epi III</td>
<td>Brown</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td><em>Chl. bathyomarinum</em></td>
<td>Green</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td><em>Chl. limicola</em> f. thiosulfatophilum</td>
<td>Green</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td><em>Chl. vibrioforme</em> strain 6030</td>
<td>Green</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td><em>Chl. vibrioforme</em> strain 1930</td>
<td>Green</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td><em>Ptc. aestuarii</em></td>
<td>Green</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>Dagowsee</td>
<td>Green</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>E3P1</td>
<td>Green</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>Fjord</td>
<td>Green</td>
<td>no</td>
<td>no</td>
</tr>
</tbody>
</table>
Figure 6.1 from Alexander et al. 2002.
Figure 6.2

Bacteriochlorophyll c

Bacteriochlorophyll e
Figure 6.3

A. pHA-nBE-aadA-HB 6747bp

B. pHA-PXE-aadA-HB 7819bp

C. pCFTXN-nBEDH-aadA 8451bp
Figure 6.4

A.

B.

- HP with homologs in cyanobacteria
- HP with homologs in archaea
- HP with homologs in other bacteria
- NTPases
- DNA modification enzymes
- Transposases
- Other
- Unique
Figure 6.5
Figure 6.6
Figure 6.7
Figure 6.8
Figure 6.9

A.

B.

C.

D.

$A_{667\text{ nm}}$

time (min.)
Figure 6.10

A.

B.

C.

D.

$A_{667 \text{ nm}}$

time (min.)
APPENDIX A

The biochemical basis for structural asymmetry in carotenoids

A.1 INTRODUCTION

In order to use light, phototrophs must absorb photons and convert the excitation energy of the photon to chemical energy. Two broad classes of phototrophs exist, and both require carotenoids. In “retinalophototrophy” (Bryant and Frigaard, 2006; Gomez Maqueo Chew and Bryant, 2007), the β-carotene cleavage product retinal absorbs light; light-induced isomerization of the retinal induces a conformational change in the rhodopsin protein to which the retinal is covalently bound (see Lanyi, 2005). The conformational change leads to directional release and ultimately the translocation of either protons or chloride ions. The second type of phototrophy, “chlorophototrophy” (Bryant and Frigaard, 2006; Gomez Maqueo Chew and Bryant, 2007) uses (bacterio)chlorophyll ((B)Chl) to harvest light. Chlorophototrophy requires not only synthesis of (B)Chl and carotenoids, which are integral to all known photosynthetic reaction centers (RCs) but also assembly of large pigment-protein complexes which orient the carotenoids and (B)Chls correctly in relation to each other and to the other cofactors. Chlorophototrophs are found in the eubacterial phyla Cyanobacteria, Proteobacteria, Chlorobi, Chloroflexi, Acidobacteria, and Firmicutes (Bryant and Frigaard, 2006; Bryant et al., submitted) and retinalophototrophs occur in haloarchaea, the Cytophagales, and a variety of marine proteobacteria, flavobacteria, and euryarchaeota. Thus, carotenoids are essential components of phototrophy and light-sensing in all three domains of life.

Chlorophototrophs synthesize two types of pigments: (bacterio)chlorophylls ((B)Chls) and carotenoids, which primarily act as photoprotective pigments but can also function in light-harvesting (e.g. Balashov et al., 2005). Although only 12 types of (B)Chls have been reported in bacteria (Gomez Maqueo Chew and Bryant, 2007), over 600 carotenoid structures have been identified (Britton et al., 2004). In retinalophototrophs, the products of oxidative cleavage of carotenoids sense light (for recent reviews on prokaryotic rhodopsins, see Lanyi, 2006; Sharma et al., 2006; Spudich, 2006; Spudich et al., 2000). Carotenogenesis, the biosynthesis of carotenoids, is an interesting case in which nature converts small, symmetric molecules into larger asymmetric molecules, many of which have specific functions in light-harvesting or light-sensing, in protection from light and oxidative stress (Olson and Krinsky, 1995), and in membrane structure and fluidity (Varkonyi et al., 2002). Because the early biosynthetic steps in this pathway involve symmetric modifications to both (identical) ends of the carotenoid, it is interesting that many of
the final products of bacterial carotenogenesis are asymmetric. Here, enzymes that can generate symmetric and asymmetric products in phototrophs are reviewed, and possible mechanisms for generating asymmetry are explored in the context of recent advances in understanding the pathways for synthesis of carotenoids in chlorophototrophs. Other recent reviews in this field concern identification of genes required for carotenogenesis (Cheng, 2006), pathway evolution (Phadwal, 2005), and utilization of known carotenoid biosynthesis proteins for efficient production of carotenoids in engineered systems (Mijts et al., 2004).

A.2 Synthesis of carotenoid precursors.

The immediate precursor of carotenoids, geranylgeranyl pyrophosphate (GGPP) can be synthesized by two pathways, the mevalonate pathway and the non-mevalonate (2C-methyl-D-erythritol 4-phosphate) pathway. Because these two pathways have been extensively reviewed (see Dubey et al., 2003; Eisenreich et al., 2004; Kuzuyama 2002; Liu et al., 2005; Qureshi and Porter, 1981; Rohdich et al., 2005), they will not be discussed in detail here. The products of both pathways are the five-carbon compounds isopentenyl diphosphate (IPP) and its isomer dimethylallyl diphosphate (DMAPP), which can be interconverted by IPP isomerase and which are the building blocks for longer isoprene compounds.

The non-mevalonate pathway is found in Chlorobium (Chl.) tepidum (Eisen et al., 2002) and other green sulfur bacteria, proteobacteria and cyanobacteria (reviewed in Eisenreich et al., 2004), and plant plastids (reviewed in Dubey et al., 2003). The mevalonate pathway is found in the plant cytoplasm (reviewed in Dubey et al., 2003), haloarchaea (Qureshi and Porter, 1981), and Chloroflexus (Cfx.) aurantiacus (Rieder et al., 1998). The pathway of terpenoid biosynthesis in heliobacteria has not been studied to date. It appears that lateral gene transfer of these two pathways has occurred frequently among diverse bacterial taxa (Boucher and Doolittle, 2000).

A.3 Phytoene, the symmetric precursor for C40 carotenoids.

In all species that synthesize C40 carotenoids, the first committed step in carotenogenesis is the head-to-head condensation of two molecules of geranylgeranyl pyrophosphate (GGPP) to generate a symmetric product, phytoene. Although in some cases only one isomer is observed (Chamovitz et al., 1992), in most cases both all-trans- and 15, 15’-cis-phytoene are produced.
Phytoene is the only carotenoid intermediate whose cis-configuration is more stable than the all-trans form (Armstrong, 1997). The 15, 15’-cis-bond is isomerized during the process of desaturation. However, it is unclear if isomerization precedes substrate binding or is caused by substrate binding to phytoene desaturase (Fraser et al., 1992; Armstrong and Hearst, 1996).

A.4 Desaturation reactions: establishing asymmetry in linear carotenoids.

In several proteobacterial species, the Chloroflexi, and in some cyanobacteria, CrtI performs four desaturations of the 7,8-, 7’,8’-, 11,12- and 11’,12’ bonds to produce a symmetric molecule, all-trans-lycopene (Linden et al., 1991; Harada et al., 2001; Giraud et al., 2004; Steiger et al., 2005; Tuschiya et al., 2005). In these species, asymmetry in carotenoids, when it occurs, is generated later in the pathway. The presence of ζ-carotene as an intermediate or minor product when CrtI from Erwinia spp., Rhodobacter (Rhb.) capsulatus, or Gloecobacter (Gl.) violaceus is expressed in E. coli suggests that CrtI performs successive desaturations on opposite sides of the molecule (Linden et al., 1991; Harada et al., 2001; Steiger et al., 2005). Based on inhibition studies in which diphenylamine was used to inhibit carotenoid biosynthesis, it is likely that CrtI from Thi-orthodospira sibirica functions in the same way (Moskalenko et al., 2005). The same type of experiments in Rhodospirillum (Rsp.) rubrum, Rhodomicrobium (Rhm.) vanniellii, Rhodopseudomonas (Rps.) globiformis, and Erythrobacter (Erb.) longus revealed 7,8,11,12-tetrahydrolycopene (asymmetric ζ-carotene) as an intermediate. This implies that the phytoene desaturases of these species may act first on one side of phytoene and then on the other (Davies, 1970; Britton et al., 1975; Leutwiler and Chapman, 1978; Schmidt and Liaaen-Jensen, 1973; Takaichi et al., 1990). Although the final product of CrtI in these species is the symmetric molecule all-trans-lycopene, the mechanistic differences exhibited by this type of CrtI enzyme might provide a way to generate carotenoid asymmetry in engineered (recombinant) systems.

In Rhb. capsulatus, Rhb. sphaeroides, and other species, CrtI performs three successive desaturations to produce neurosporene (Giuliani et al., 1986; Armstrong et al., 1990; Linden et al., 1991; Lang et al., 1995; Harada et al., 2001). This type of CrtI enzyme produces an asymmetric product, neurosporene, which can be converted into a variety of end products in various organisms. The final products spheroidene, spheroidenone, and their derivatives have asymmetric chromophores, and the ψ- and 7,8-dihydro-ψ ends of these molecules are modified differently.
(Takaichi, 1999). CrtI from *Rhb. capsulatus* has been studied extensively *in vitro*. This enzyme has much higher activity on the symmetric intermediates phytoene and ζ-carotene than on the asymmetric intermediate phytofluene (Raisig *et al*., 1996). The enzyme desaturates the 11,12, 11′,12′ and 7,8 bonds in phytoene and is also able to desaturate the 7,8 bonds in diapophytoene and diapo-ζ-carotene (Raisig & Sandmann, 2001). The appearance of ζ-carotene as an intermediate (Giuliano *et al*., 1986; Raisig *et al*., 1996; Raisig & Sandmann, 2001) suggests that, similarly to CrtI of *Erwinia* spp., this CrtI first desaturates the 11,12 and 11′,12′ bonds and then desaturates the 7,8 bond. Because of its activity on diapocarotenoids, it has been suggested that *Rhb. capsulatus* CrtI recognizes as a substrate any single bond within a potentially conjugated system of three or more bonds on one half of the substrate but only a single double bond on the other (Raisig & Sandmann, 2001). However, if this were the case, the enzyme should catalyze at least four successive desaturations. Instead, this CrtI enzyme is the earliest known enzyme in carotenogenesis that can generate an asymmetric product.

Cyanobacteria and GSB also produce lycopene before producing asymmetric carotenoids. However, in these organisms CrtP desaturates phytoene to ζ-carotene, which is usually a mixture of all-*trans* species as well as different *cis*-isomers (Linden *et al*., 1991; Fraser *et al*., 1993; Schneider *et al*., 1997; Frigaard *et al*., 2004; Bautista *et al*., 2005). CrtQ then desaturates ζ-carotene twice more to produce 7,9,7′,9′-*tetra-cis*-lycopene (prolycopene) and other lycopene isomers (Albrecht *et al*., 1996; Breitenbach *et al*., 1998; Frigaard *et al*., 2004). The lycopene isomers are subsequently converted to all-*trans*-lycopene by the carotenoid isomerase CrtH (Breitenbach *et al*., 2001; Masamoto *et al*., 2001; Frigaard *et al*., 2004). Although all *Chlorobi* and cyanobacteria, including *G. violaceus*, appear to have CrtH, no other phototrophic bacteria has yet been shown to have such an enzyme.

The occurrence of this three-enzyme (CrtP, CrtQ, CrtH), multi-intermediate pathway in place of the single-enzyme process mediated by CrtI is curious, and no completely satisfactory explanation for its occurrence has yet been offered. One possibility is that this multi-step pathway is required because of the presence of *cis*-carotenoids in Type I reaction centers—a requirement that may also necessitate the production of highly specific *cis*-isomers. Both cyanobacterial photosystem I (Jordan *et al*., 2001) and the reaction centers of green sulfur bacteria (Bi-alek-Bylka *et al*., 1998) contain *cis*-carotenoids. Because *cis*-*trans* isomerization of colored carotenoids can be induced by light, the existence of the carotenoid isomerase, CrtH, is intriguing.
Experiments with *Scenedesmus obliquus* as well as the *crtH* mutant of *Chl. tepidum* demonstrate that light-activated carotenoid isomerization is sufficient for photoautotrophic growth of these organisms (Sandmann 1991; Frigaard *et al.*, 2004). Characterization of a *crtH* mutant in *Synechocystis* sp. PCC 6803 showed that all-trans β-carotene is required for PSII assembly and that PS II is not assembled in the absence of both CrtH and light (Masamoto *et al.*, 2004). Other enzymes in carotenoid biosynthesis, such as the plant 9-*cis*-epoxy dioxygenase (NCED) family, only act on specific *cis*-isomers (Schwartz *et al.*, 1997). Thus, while it seems likely that CrtH produces all-trans-lycopene from other lycopene isomers, it is nevertheless also possible that CrtH may be used to produce specific *cis*-isomers of lycopene.

### A.5 ψ-end modifications

In the purple bacteria, carotenoids are only rarely cyclized. More often, the ψ-ends are hydroxylated and/or ketolated, and additional functional groups may be linked to the hydroxyl group. In many species three enzymes act on these linear end groups: CrtC hydrates the 1,2 and/or 1’,2’ double bond, leaving a hydroxyl group on the 1 and/or 1’ carbon (Scolnik *et al.*, 1980); CrtD desaturates the 3,4 and/or 3’,4’ bond (Albrecht *et al.*, 1997); and CrtF methylates the 1- and/or 1’-hydroxyl group (Scolnik *et al.*, 1980). In species that produce spheroidene and spheroidenone, these three enzymes act on a ψ-end of neurosporene; in species that produce spirilloxanthin, these three enzymes first collectively modify a single ψ-end of lycopene and then modify the other end (Pinta *et al.*, 2003). However, the asymmetry of neurosporene is not the only reason that hydration occurs on only one end of this molecule. CrtC from *Rhb. capsulatus*, which synthesizes both spheroidene and spheroidenone, recognizes 1-hydroxylycopene as a substrate, but not 1-hydroxy-3,4 didehydrolycopene (Steiger *et al.*, 2002). Similarly, CrtD desaturates 1-hydroxylycopene, but not 1,1’-dihydroxylycopene (Steiger *et al.*, 2000). These results indicate that even though CrtC and CrtD act on the extreme ends of their substrates, they recognize the entire molecule and not just the half to be modified.

Ketolation of the 2- and/or 2’ carbon to produce spheroidenone or diketospirilloxanthin occurs by the action of CrtA (Lang *et al.*, 1995), which appears to act after the methoxy group has been added to the 1- or 1’ position (Takaichi and Shimada, 1999; Pinta *et al.*, 2003). Since CrtA acts only after CrtC, CrtD, and CrtF have modified the ψ-end(s) of neurosporene or lypo-
pene, substrate recognition by CrtA probably requires the presence of the 1- or 1’-methoxy group. Thus, CrtA probably is not responsible for introduction of asymmetry.

Although cyanobacteria also produce 1 and/or 1’ hydroxylated carotenoids, including myxol, oscillol, and myxoxanthophylls (Takaichi et al., 2006; Tsuchiya et al., 2005; Aakermann et al., 1992; Schaper et al., 2005; Schagerl and Donebaum, 2003), no cyanobacterial genome encodes a homolog of crtC. Similarly, Chloroflexi also produce 1’-hydroxylated carotenoids, which are usually glycosylated (see below), but genomic sequences of these organisms likewise do not contain crtC homologs. Instead, cyanobacteria, Cfx. aurantiacus and Roseiflexus sp. have a carotenoid hydroxylase, CruF, that is related to an open reading frame (ORF) found in the carotenogenesis gene cluster of the marine, myxol-producing, marine bacterium P99-3 (J. E. Graham and D. A. Bryant, unpublished results). When cruF from the cyanobacterium Synechococcus sp. PCC 7002 is expressed an E. coli strain that produces lycopene, both 1-hydroxy- and 1,1’-di-hydroxylycopene are produced. To our best knowledge, the gene encoding the 2’ ψ-end hydroxylase required for myxol synthesis in cyanobacteria has not been identified.

The hydroxylated ψ-ends of carotenoids in Chlorobi and Chloroflexi (Takaichi et al., 1997), as well as myxoxanthophyll-producing cyanobacteria (Takaichi et al., 2001, 2005), are not methylated but are glycosylated. The glycosyltransferase CruC of GSB was recently identified by gene neighborhood comparisons with non-photosynthetic carotenogenic bacteria that produce similar carotenoids (Maresca and Bryant, 2006). CruC is unrelated to the ring glucosyltransferase CrtX found in Pantoea (formerly Erwinia) spp. The genomes of Chloroflexi and myxoxanthophyll-producing cyanobacteria do not contain genes homologous to crtX. However, insertional inactivation of a gene encoding a glycosyltransferase homologous to the product of ORF sll1004 of Synechocystis sp. PCC 6803, denoted CruG, results in complete loss of myxoxanthophyll synthesis in the unicellular marine cyanobacterium Synechococcus sp. PCC 7002 (J. E. Graham and D. A. Bryant, unpublished results). Although this glycosyltransferase is expected to modify the 2’-hydroxyl group rather than a 1’ hydroxyl group, CruG has significant sequence similarity to CruC of GSB. The carotenoid acyltransferase, CruD, from GSB was identified in the same way and has homologs in the Chloroflexi and non-photosynthetic carotenogenic bacteria (Maresca and Bryant 2006). Although it is tempting to speculate that competition between CrtC and CruA, the lycopene cyclase (see below), prevents cyclization of the second ψ-end in green-colored GSB, OH-chlorobactene glucoside laurate accounts for only ~10% of the
total carotenoid content in GSB, and therefore such competition probably does not account for
the carotenoid asymmetry observed in these organisms. In myxoxanthophyll-producing cyanobacteria, control of substrate fluxes between the branched pathways leading to β-carotene and its
derivatives and to myxoxanthophyll may result from competition between CruA, the major carotenoid cyclase, and CruF, the 1’ hydroxylase (J. E. Graham, and D. A. Bryant, unpublished re-

tsults).

A.6 Cyclization reactions

In green-colored GSB, *Chloroflexi*, and cyanobacteria, asymmetry in carotenoids is established
by lycopene cyclization. Four classes of lycopene cyclases have been identified in bacteria. CrtY-type β-cyclases are found in *Streptomyces* spp. (Krugel et al. 1999), the *Chloroflexi*, and
some carotenogenic proteobacteria (e.g. Misawa et al., 1990; Matsumura et al. 1997). Although heterodimeric cyclases have not yet been found in any chlorophototroph, such cyclases are found in some Gram-positive bacteria (Krubasik, 2000a; Viveiros, 2000), and are related to the lycopene cyclases of the archaea *Halobacterium salinarum* (Peck et al. 2002) and *Sulfolobus solfatarius* (Hemmi et al., 2003). The CrtL family includes the β- and ε-cyclases in some cyanobac-
teria and plants (Cunningham et al., 1994; Stickforth et al. 2003). Finally, the CruA family of
lycopene cyclases is found in *Chlorobi*, cyanobacteria lacking CrtL, and plants (Maresca et al.,
2004). These four classes are distantly related to each other and share a few conserved motifs: all except the heterodimeric type share an N-terminal flavin-binding domain (Krubasik and Sand-
mann, 2000b; Maresca et al., 2004). Enzymes that produce monocyclic carotenoids are present
in the CruA-, CrtY- and CrtL-type classes (Maresca et al., 2004; Cunningham and Gantt, 1994; Krubasik and Sandmann, 2000b; Tao et al. 2004; Teramoto et al., 2003), and there are no obvi-
ous sequence differences that distinguish lycopene monocyclases from dicyclases (Cunningham
and Gantz, 1994).

In green-colored GSB, which produce the monocyclic aromatic chlorobactene, the first
asymmetric intermediate is γ-carotene (Maresca et al., 2004). When it is expressed in a lycopene-
producing strain of *E. coli*, the lycopene cyclase of *Chl. tepidum*, CruA, converts up to 30% of
the available lycopene to β-carotene. However, in *Chl. tepidum*, neither β-carotene nor its aro-
matic derivative isorenieratene has ever been observed (Takaichi et al., 1997; Frigaard et al.,
2004). This suggests that, although γ-carotene is the first observed asymmetric intermediate in
this pathway, some process other than substrate recognition must prevent cyclization of the ψ-end of γ-carotene. One intriguing possibility is that CrtH, in the process of isomerizing prolycopene to all-trans-lycopene, leaves one cis-bond, and the substrate for CruA is, in fact, a mono-cis-lycopene isomer.

In many species, including brown-colored species of GSB, cyclization is a branch point between the monocyclic and dicyclic pathways. Brown-colored species of GSB produce primarily isorenieratene and β-isorenieratene, both of which are aromatic derivatives of β-carotene (Liaaen-Jensen, 1965; Hirabayashi et al., 2004). Additionally, a small amount of acylated hydroxychlorobactene glycoside is produced (Liaaen-Jensen, 1965; Hirabayashi et al., 2004). The genome sequences of three brown-colored species of GSB reveal that these genomes encode two copies of cruA. One gene (cruA) encodes a lycopene monocyclus, and the paralogous gene, denoted cruB, is a γ-carotene cyclase (Maresca, 2007). These observations establish a highly counter-intuitive situation. Brown-colored GSB presumably produce a symmetric substrate, all-trans-lycopene, and they have two apparently different enzymes that act on opposite ends of this substrate molecule to produce a symmetric product, β-carotene. In this case, competition between the γ-carotene cyclase and the ψ-end hydroxylase (CrtC) could control the yield of the two pathways—one leading to the acylated monocyclic carotenoids and one leading to the dicyclic carotenoids that are principally transported into the chlorosome. Alternatively, one of these enzymes is a lycopene monocyclus with low activity, which directs its products into the chlorobactene pathway, and the other is a dicyclus and is much more active.

It is also possible that the branch point occurs prior to the cyclization step. Accumulation of all-trans-neurosporene and some cis-neurosporene isomers has been noted in several organisms with the CrtP-CrtQ-CrtH pathway preceding cyclization (Sandmann, 1991; Masamoto et al., 2001; Hirabayashi et al., 2004), and it has been suggested that CrtH, in addition to isomerizing prolycopene to all-trans-lycopene, converts earlier intermediates into suitable substrates for CrtQ (Masamoto et al., 2001). Studies with inhibitors of lycopene cyclase activity in purple bacteria have also shown accumulation of neurosporene when cyclization is blocked (Leutwiler and Chapman, 1978). Thus, it seems possible that, in some species at least, branching of the pathway either between mono- and dicyclic carotenoids or between dicyclic and linear carotenoids may occur when neurosporene, rather than lycopene, is isomerized. It is interesting to note that lycopene accumulates when cyclization is blocked in Chl. tepidum (Chlorobi), which normally syn-
thesizes the monocyclic, aromatic carotenoid, chlorobactene (Takaichi et al., 1997; Frigaard et al., 2004).

The Chloroflexi produce derivatives both of g-carotene and β-carotene (Hanada et al., 2001; Takaichi et al., 2001; Gich et al., 2003). Although no biochemical or genetic work has been done with any of these species to characterize carotenoid biosynthesis, as a group they appear to present an interesting case of biosynthetic pathways which seem to be cobbled together from what appear to be several horizontal gene transfer events from different species. The draft genome sequences of the species Chloroflexus (Cfx.) aurantiacus, Cfx. aggregans, Roseiflexus sp. strain RS1, Roseiflexus castenholzii, and Herpetosiphon (Hps.) aurantiacus are all available at NCBI (http://www.ncbi.nlm.nih.gov). Based on BLAST searches (Altschul et al., 1997) with a cutoff e-value of less than 10^{-20}, all species except Cfx. aggregans have homologs of CrtB and all have two or three homologs of CrtI. The two Chloroflexus spp. have homologs of CrtP from the cyanobacterium Synechocystis sp. PCC 6803 and lycopene cyclases in the CrtY family, while the two Roseiflexus spp. have CrtL-type cyclases, and Hps. aurantiacus has a CruA-type cyclase. These Chloroflexi as well as Heliothrix oregonensis make carotenoid glycoside esters (Takaichi et al., 1997, 2001; Hanada et al., 1995, 2002; Pierson et al., 1985; Schmidt, 1980) but do not contain homologs to CrtX. However, all Chloroflexi genomes encode homologs of the CruC-type glycosyltransferase as well as homologs of the carotenoid acyltransferase CruD (Maresca and Bryant, 2006). Rfx. castenholzii, Cfx. aurantiacus and Heliothrix oregonensis also synthesize carotenoids with ketolated rings (Pierson et al., 1985; Schmidt, 1980; Takaichi et al., 2001). The carotenoid intermediates observed under oxic conditions include neurosporene and asymmetric ζ-carotene, but no lycopene; this observation suggests that the asymmetry in this biosynthetic pathway may be established with the desaturation of phytoene by CrtI, rather than by monocyclization of lycopene (Takaichi et al., 2001). Other species of filamentous anoxygenic phototrophs (FAPs) also make carotenoids under oxic conditions (Hanada et al., 1995, 2002; Holt & Lewin, 1968), and detailed analyses of other members of this group might reveal more widespread production of either echinenone or canthaxanthin.

The primary carotenoid in cyanobacteria is β-carotene, but the γ-carotene derivatives myxol and myxoxanthophyll are common in these species (Santoyo et al., 2006; Schagerl and Muller 2006; Takaichi et al., 2006, 2005; Schluter et al., 2004; Lakatos et al, 2001; Hesse et al., 2001; Miskiewicz et al., 2000; Mohamed and Vermaas, 2004). In all of these cases, some
mechanism for ensuring production of an appropriate ratio of mono- to di-cyclic carotenoids must exist. Cyanobacteria that do not have a CrtL-type lycopene cyclase instead have two cyclases of the CruA subfamily, denoted CruA and CruP (Maresca et al., 2004; Maresca, Graham, et al., in preparation). It seems logical to postulate that one cyclase is a lycopene monocyclase that directs its products into the myxol/myxoxanthophyll pathway and that one is a lycopene di-cyclase, which directs its products into the β-carotene/echinenone/zeaxanthin pathway. Consistent with this hypothesis, expression of CruP in an E. coli strain that produces lycopene leads to γ-carotene synthesis, and insertional inactivation of cruA produces a strain that accumulates lycopene (J. E. Graham and D. A. Bryant, unpublished results). Alternatively, competition between the cyclases and the 1' hydroxylase could determine which pathway a molecule enters; 1' hydroxylation would presumably be followed by 3',4' desaturation, 2' hydroxylation, and finally the attachment of a sugar moiety to the 2' hydroxyl, as proposed by Takaichi and colleagues (2004, 2005). Although this pathway can be predicted, as noted above, cyanobacteria lack identifiable homologs of crtC and instead have a 1'-hydroxylase, encoded by cruF, which has been identified though interposon mutagenesis in Synechococcus sp PCC 7002 and by overexpression in E. coli. Although the glycosyltransferase, CruG, has also been identified, the 2'-hydroxylase has not yet been identified in cyanobacteria (see above). Mohamed and Vermaas (2004) reported that Synechocystis sp. PCC 6803 ORF slt1293 encodes the 3',4' desaturase, CrtD. However, this result has not yet been independently validated. The apparently homologous protein in Synechococcus sp. PCC 7002 appears to have a different function and does not affect myxoxanthophyll biosynthesis as predicted (J. E. Graham and D. A. Bryant, unpublished data).

A.7 Ring modifications

In Chlorobi and actinomycetes, carotenoids with aromatic end groups are produced by the γ-carotene desaturase, CrtU, which desaturates the ring and transfers one of the 1-methyl groups to the 2-position on the ring (Frigaard et al., 2004; Krugel et al., 1999; Viveiros et al., 2000). The CrtU gene of phototrophs is interesting because of its bifunctional, multidomain properties. CrtU has a Rieske FeS cluster domain that is inserted between a duplication of the flavin/carotenoid binding domain of phytoene dehydrogenases. CrtU homologs have also been found in the genomes of some but not all cyanobacteria (Krugel et al., 1999; Frigaard et al., 2004; Mohamed and Vermaas, 2006). It has generally been assumed that cyanobacteria do not synthesize aro-
matic carotenoids (e.g., Krugel et al., 1999; Mohamed and Vermaas, 2006), and this assumption has led some to conclude that cyanobacterial CrtU must have a different function than the CrtU homologs in GSB and actinomycetes. Mohamed and Vermaas (2006) tentatively assigned the CrtU ortholog of *Synechocystis* sp. PCC 6803, Sll0254, the function of lycopene cyclase/dihydroxylase. Analysis of the *crtU* ortholog in *Synechococcus* sp. PCC 7002 indicates that it desaturates the rings of β-carotene and transfers the two 1-methyl groups to the 3- and 4-positions of the rings (J. E. Graham and D. A. Bryant, unpublished results). The resulting purpurin, renierapurpurin, is subsequently oxidized, through the participation of a di-iron oxygenase denoted CruH, to produce a novel, dicarboxylate carotenoid: χ,χ-caroten-18, 18'-dioic acid ("synechoxanthin"). It is presently not known if CruH can fully oxidize the 4- and 4'-methyl groups to the corresponding carboxylic acids or if a dehydrogenase might also be required.

In many species that produce β-carotene, a significant fraction of the β-carotene is modified by hydroxylation and/or ketolation of the rings to produce cryptoxanthin, zeaxanthin, echinenone, hydroxyechinenone, canthaxanthin, or astaxanthin (e.g. Misawa et al., 1995; Balashov et al., 2006; Tao and Cheng, 2004; Ye et al., 2006; Mochimaru et al., 2005). In most cases, these enzymes act on any available rings: in both *Synechocystis* sp. PCC 6803 CrtR is responsible for the introduction of 3- and 3'-hydroxyl groups onto the rings of the dicyclic zeaxanthin (Masamoto et al., 1998) as well as the monocyclic myxoxanthophyll (Lagarde and Vermaas, 1999). The same results have been found in *Synechococcus* sp. PCC 7002 (J. E. Graham and D. A. Bryant, unpublished results). The most likely explanation for this is that CrtR recognizes only the end of the carotenoid with the β-ring, and disregards the structure of the other end of the molecule. In contrast, most cyanobacterial ketolases act asymmetrically on only one end of the substrate. Cyanobacteria have been shown to possess either CrtO-type ketolases or CrtW-type ketolases. In some species β-carotene is ketolated by CrtO only at the 4-position to produce echinenone (Fernandez-Gonzalez et al., 1997; Mochimaru et al, 2005). CrtW-type enzymes from *Nostoc punctiforme* have been reported to convert β-carotene to canthaxanthin via echinenone (Steiger and Sandmann, 2004). When *crtO* or *crtW* are inactivated in *Nostoc* sp. PCC 7120, neither strain makes canthaxanthin (Mochimaru et al., 2005). The *crtO* mutant strain additionally does not synthesize echinenone, while the *crtW* mutant strain does not synthesize 4-ketomyxol-2-fucoside (Mochimaru et al., 2005). In this case, CrtW recognizes only those β-rings associated with myxol, while CrtO recognizes only β-carotene as a substrate. In *Synechococcus* sp. PCC
7002, which only has a CrtW-type ketolase, inactivation of the \textit{crtW} gene principally led to loss of echinenone (J. E. Graham and D. A. Bryant, unpublished results). Since \textit{Synechococcus} sp. PCC 7002 does not accumulate significant amounts of canthaxanthin but does produce hydroxyechinenone, it is possible that CrtW can act symmetrically but is out-competed by the CrtR 3’-hydroxylase.

A.8 C30 backbones: Synthesis of apocarotenoids in Heliobacteria

Very little has been published on the carotenoids of the Heliobacteria, and even less on the biosynthesis of those carotenoids. Heliobacteria make only apocarotenoids, with the primary product 4, 4’-diaponeurosporene (Takaichi \textit{et al.}, 1997a). Because small amounts of the C30 precursors 4, 4’-diapo-ζ-carotene, 4, 4’-diapophytofluene, and 4, 4’-diapophytoene accumulate in the cells (Takaichi \textit{et al.}, 1997a), these carotenoids are most likely synthesized by a C30 pathway similar to that for staphyloxanthin in \textit{Staphylococcus aureus} (Wieland \textit{et al.}, 1994) rather than by a C40 pathway and subsequent cleavage, as in plants (reviewed in Auldridge \textit{et al.}, 2006). The first steps in staphyloxanthin (α-D-glucopyranosyl-1-O-(4, 4’-diaponeurospren-4-oate) 6-O-methyltetradecanoate) biosynthesis are the condensation of two molecules of farnesyl phyrophosphate by the enzyme CrtM to produce 4, 4’-diapo-phypoene, followed by the desaturation of dehydrosqualene to 4,4’-diaponeurosporene (Wieland \textit{et al.}, 1994). Although no homolog to CrtM can be detected in the \textit{Heliobacillus (Hbc.) mobilis} genome, this genome is still in draft form and such a gene may appear when the genome has been completed.

The 4, 4’-diapo-phypoene desaturase CrtN, like CrtI of \textit{Rhodobacter} \textit{spp.}, performs three successive desaturations to produce an asymmetric molecule (Wieland \textit{et al.}, 1994). An ORF homologous to CrtN is found in the photosynthesis gene cluster in \textit{Hbc. mobilis} (Xiong \textit{et al.}, 1998), and most likely encodes a three-step 4,4’-diapo-phypoene desaturase. The appearance of 4, 4’-diapo-ζ-carotene rather than 4, 4’-diapo-asymmetric-ζ-carotene in wild-type \textit{Hbc. mobilis} and other heliobacterial species (Takaichi \textit{et al.}, 1997a) implies that CrtN desaturates first the 11,12 and 11’,12’ bonds and then desaturates the 7,8 bond, again similar to CrtI of \textit{Rhodobacter} \textit{spp.} However, direct genetic and biochemical evidence for the activity of CrtN in heliobacteria as well as identification of a 4,4’-diapophytoene (squalene) synthase are still necessary.
A.9 Oxidative cleavage of carotenoids

All known rhodopsins contain a retinal chromophore that is covalently bound to the protein (Spudich 2000). In archaea and marine proteobacteria, retinal is synthesized by oxidative cleavage of β-carotene at the 15, 15’ bond to produce two identical molecules of retinal (Martinez et al., 2007; Peck et al., 2001; Sabehi et al., 2005). The recently-described apocarotenoid oxygenases (SynACO and NosACO) of cyanobacteria only cleave apo-carotenoids to produce retinal (Marasco et al., 2006; Ruch et al., 2005; Scherzinger et al., 2006). Based on the crystal structure of SynACO, the apocarotenoid oxygenase from Synechocystis sp. PCC 6803, only linear end groups can enter the substrate channel (Kloer et al., 2005). In fact, only all-trans linear end groups can enter the substrate channel, but as they enter, interactions with the protein force the 13, 14 and 13’, 14’ bonds into cis-conformations (Kloer et al., 2005).

The question of the precursor of the apocarotenoid substrate in cyanobacteria remains open. Although it seems unlikely that there is a separate pathway for synthesis of apocarotenoids in cyanobacteria, similar to that found in Staphylococcus and the Heliobacteria, several cyanobacterial genomes encode several homologs to crtI, but these ORFs have no characterized function. If, on the other hand, the apocarotenoids are produced by cleavage of a C40 compound, it will be important to establish whether the substrate for cleavage is γ-carotene, its derivatives, or β-carotene.

A.10 Conclusions

Although the core set of enzymes required for carotenoid biosynthesis is very similar in most organisms, these enzymes vary sufficiently in their substrate specificities and products that much of the carotenoid diversity is actually generated early in the pathway. Related proteins that can produce slightly different products translates to differences in end products: neurosporene or lycopene, γ- or β-carotene, echinenone or canthaxanthin. Future work in this field should investigate the structural differences between these pairs of enzymes as well as the precise structural differences for their substrates.
References


Figure Legends.

Figure 1. Synthesis of phytoene, the first committed step in carotenogenesis, by head-to-head condensation of two molecules of geranylgeranyl pyrophosphate by phytoene synthase, CrtB. Phytoene is one of the few carotenoids whose 15,15'-cis-isomer is more stable than the all-trans form.

Figure 2. A. Desaturation of phytoene to either neurosporene or lycopene by CrtI. The different types of CrtI and the order in which the bonds are oxidized are discussed in section A.4. B. Desaturation of phytoene in green sulfur bacteria and cyanobacteria by CrtP and CrtQ to tetra-cis-lycopene, and subsequent isomerization by CrtH. This pathway is also found in plants.

Figure 3. A. Modification of the ψ end of neurosporene to spheroidene and spheroidenone. B. Modification of the two ψ ends of lycopene to produce spirilloxanthin or diketospirilloxanthin. The same set of enzymes acts in both the spheroidene/spheroidenone and spirilloxanthin/diketospirilloxanthin pathways, but does not seem to recognize ψ end groups with an adjacent 7,8 or 7′,8′ bond that is fully reduced.

Figure 4. A. Modification of the ψ end of γ-carotene in cyanobacteria to produce myxoxanthophyll. The 2′ hydroxylase has not yet been identified. B. Modification of the ψ end of chlorobactene in green sulfur bacteria. The product, OH-chlorobactene glucoside laurate, has been identified in most species of green sulfur bacteria, and the genes encoding the glycosyltransferase and acyltransferase, CruC and CruD, have been found in all available green sulfur bacterial genomes.

Figure 5. Isomerization of carotenoid ψ end group to either ε- or β-ionone rings. Although ε-cyclases are only found in the CrtL family of lycopene cyclases, β-cyclases are found in all 4 families.

Figure 6. Synthesis of cyclic carotenoids from lycopene in green sulfur bacteria. The green-colored species produce the monocyclic carotenoid γ-carotene and its aromatic derivative chlorobactene; brown-colored species produce the dicyclic β-carotene and its aromatic derivatives β-isorenieratene and isorenieratene.

Figure 7. Cyclization of lycopene by CruA and CruP in cyanobacteria. CruP has been shown by heterologous expression in E. coli to be a lycopene monocyclase; based on insertional inactivation of cruA in Synechococcus sp. PCC 7002, CruA appears to be the major carotenoid cyclase in the cells.

Figure 8. β-carotene derivatives in cyanobacteria.

Figure 9. Synthesis of apocarotenoids in Staphylococcus aureus. The pathway may be similar in the phototrophic Heliobacteriaceae. A homolog of CrtN has been found in a photosynthetic gene cluster in Heliobacillus mobilis; however, no homolog to crtM has yet been found in the genome.
Figure 10. A. Symmetric oxidative cleavage of β-carotene by Blh or Brp to produce two molecules of retinal. Homologs of Blh and Brp are commonly found in the haloarchaea, marine proteobacteria, and other retinalophototrophs. B. Asymmetric oxidative cleavage of β-apo-8’-carotenol by the cyanobacterial enzymes NosACO and SynACO to produce retinal.
Figure 1.

```
geranylgeranyl pyrophosphate

CrtB

all-trans-phytoene

15, 15'-cis-phytoene
```
Figure 2.

A. 

15, 15' cis-phytoene

neurosporene

all-trans- lycopene

B. 

15, 15' cis-phytoene

9, 9' di-cis-ζ-carotene

7,9,7',9' tetra-cis-lycopene

all-trans- lycopene
Figure 3.

A

- neurosporene
  \[\xrightarrow{\text{CrtC}}\]
- 1-hydroxyneurosporene
  \[\xrightarrow{\text{CrtD}}\]
- 1-hydroxy-3,4 dihydroneurosporene
  \[\xrightarrow{\text{CrtF}}\]
- spheroidene
  \[\xrightarrow{\text{CrtA}}\]
- spheroidenone

B

- lycopene
  \[\xrightarrow{\text{CrtCDF}}\]
- 1-methoxy-lycopene
  \[\xrightarrow{\text{CrtCDF}}\]
- spirilloxanthin
  \[\xrightarrow{\text{CrtA}}\]
- diketospirilloxanthin
Figure 4A.

\[ \gamma\text{-carotene} \xrightarrow{\text{CruF}} \]

\[ 1'\text{-OH-}\gamma\text{-carotene} \xrightarrow{??} \]

\[ 1'\text{-OH-torulene} \xrightarrow{??} \]

\[ 1', 2'\text{-di-OH-torulene} \xrightarrow{\text{CrtR}} \]

\[ \text{myxol} \xrightarrow{\text{CruG}} \]

\[ \text{myxoxanthophyll} \]
Figure 4B.
Figure 6.
Figure 7.
Figure 8.
Figure 9.
Figure 10.

A

\[ \beta\text{-carotene} \xrightarrow{B_{lh}} B_{rp} \]

\( \times 2 \)

retinal

B

\[ \beta\text{-apo-8'}\text{-carotenol} \xrightarrow{\text{SynACO}} \xrightarrow{\text{NosACO}} \]

retinal

\( + \)

\[ \text{O=O} \quad \text{O=S=O} \text{OH} \]
APPENDIX B.

GENETIC MANIPULATION OF CAROTENOID BIOSYNTHESIS IN THE GREEN SULFUR BACTERIUM *CHLOROBIUM TEPIDUM*
Genetic Manipulation of Carotenoid Biosynthesis in the Green Sulfur Bacterium *Chlorobium tepidum*†

Niels-Ulrik Frigaard, Julia A. Maresca, Colleen E. Yunker, A. Daniel Jones, and Donald A. Bryant

*Department of Biochemistry and Molecular Biology and Department of Chemistry, The Pennsylvania State University, University Park, Pennsylvania 16802*

Received 16 March 2004/Accepted 14 May 2004

The green sulfur bacterium *Chlorobium tepidum* is a strict anaerobe and an obligate photoautotroph. On the basis of sequence similarity with known enzymes or sequence motifs, nine open reading frames encoding putative enzymes of carotenoid biosynthesis were identified in the genome sequence of *C. tepidum*, and all nine genes were inactivated. Analysis of the carotenoid composition in the resulting mutants allowed the genes encoding the following six enzymes to be identified: phytoene synthase (crtB/CT1386), phytoene desaturase (crtP/CT0807), γ-carotene desaturase (crtQ/CT1414), α-carotene desaturase (crtH/CT0323), carotenoid 1,2'-hydratase (crtC/CT0301), and carotenoid cis-trans isomerase (crtI/CT0649). Three mutants (CT0180, CT1357, and CT1416 mutants) did not exhibit a discernible phenotype. The carotenoid biosynthetic pathway in *C. tepidum* is similar to that in cyanobacteria and plants by converting phytoene into lycopene using two plant-like desaturases (CrtP and CrtQ) and a plant-like cis-trans isomerase (CrtI) and thus differs from the pathway known in all other bacteria. In contrast to the situation in cyanobacteria and plants, the construction of a *crtB* mutant completely lacking carotenoids demonstrates that carotenoids are not essential for photosynthetic growth of green sulfur bacteria. However, the bacteriochlorophyll *a* contents of mutants lacking colored carotenoids (*crtB, crtP, and crtQ* mutants) were decreased from that of the wild type, and these mutants exhibited a significant growth rate defect under all light intensities tested. Therefore, colored carotenoids may have both structural and photoprotection roles in green sulfur bacteria. The ability to manipulate the carotenoid composition so dramatically in *C. tepidum* offers excellent possibilities for studying the roles of carotenoids in the light-harvesting chlorosome antenna and iron-sulfur-type (photosystem I-like) reaction center. The phylogeny of carotenogenic enzymes in green sulfur bacteria and green filamentous bacteria is also discussed.

Carotenoids are synthesized by all photosynthetic organisms and by many nonphotosynthetic microorganisms (7, 12). Carotenoids are also important nutritional supplements for an even wider range of organisms, including animals and humans. In photosynthetic organisms, carotenoids function in light harvesting, photoprotection, and in some systems, structure stabilization (18). The light-harvesting function involves absorption of blue light with wavelengths from 400 to 550 nm and transfer of excitation energy to chlorophylls (Chls) or bacteriochlorophylls (BChls) via singlet states (18). Photoprotection is conferred by carotenoids by their ability to quench (B)Chl triplet states and to scavenge singlet oxygen and harmful radicals. Finally, carotenoids appear to serve a structural role in assembly and stabilization of some pigment-protein complexes in purple bacteria and plants (46).

The first committed step in the biosynthesis of C₄₀ carotenoids is the head-to-head condensation of two C₂₀ geranylgeranyl diphosphate molecules by phytoene synthase to form phytoene (7, 12). Phytoene is then converted to all-trans-lycopene by one of two principal pathways: (i) desaturation and cis-trans isomerization by a single enzyme, CrtI (which may produce all-trans-neurosporene in some microorganisms), or (ii) desaturation to cis-γ-carotene by CrtP, desaturation of cis-γ-carotene to cis-lycopene by CrtQ, and finally cis-trans isomerization by CrtH to produce all-trans-lycopene (7, 12, 30). The CrtI-dependent pathway is found in fungi and most prokaryotes, whereas the CrtP/CrtQ/CrtH-dependent pathway until now has been found only in oxygenic photosynthetic organisms (cyanobacteria, algae, and plants). Cyclization, catalyzed by a lycopene cyclase, may then take place at one or both ends of lycopene. Neurosporene, lycopene, and their cyclization products, may be modified by additional enzymes in various organisms to produce a plethora of different carotenoids (7, 12, 52).

Green sulfur bacteria are obligately photoautotrophic and strictly anaerobic, and they form a distinct phylogenetic group (29, 44). These bacteria contain very large and unique antenna complexes, called chlorosomes, whose major pigment is BChl *c, d*, or *e* (8, 9). The major and characteristic carotenoids of green sulfur bacteria are chlorobactene (*φ,ψ*-carotene) and isorenieratene (*φ,φ*-carotene), and these carotenoids are unusual by having aromatic *φ* end groups (38, 39, 52). Aromatic carotenoids with *φ* end groups are otherwise found only in some actinomycetes and a few sponges (*Reniera* spp.) (7, 12).

Because it is naturally transformable (19) and because its genome has been sequenced (15), *Chlorobium tepidum* (57)
has emerged as the principal model organism for studies of green sulfur bacteria. Much has already been learned about the biosynthesis of BChl c in this organism by bioinformatic and gene inactivation approaches (22–24), and one objective of the present study was to study the biosynthesis and functions of carotenoids in the same organism by similar approaches. The major carotenoid in C. tepidum is chlorobactene; minor carotenoids are γ-carotene, 1',2'-dihydrorhodobactene, 1',2'-dihydro-γ-carotene, OH-chlorobactene, and the laurate esters of both OH-chlorobactene glucoside and OH-γ-carotene glucoside (53). In C. tepidum, about 90% of the cellular carotenoids are located in the chlorosomes (20, 54). These carotenoids are not easily removed by detergent treatment of isolated chlorosomes (21). This observation suggests that the majority of the carotenoids are located in the chlorosome interior. The predominant carotenoids associated with isolated reaction center complexes are OH-carotenoid glucoside laurate esters; therefore, these carotenoids probably play a specific role in the reaction center (54).

By using a combination of genome analysis and gene inactivation in C. tepidum, we have elucidated the biosynthetic pathway leading to chlorobactene and OH-chlorobactene (Fig. 1). We have also established that the enzymes converting phytoene to lycopene are related to those in the lineage of oxygenic photosynthetic organisms (cyanobacteria, algae, and plants), which raises some interesting evolutionary issues. The observation that carotenoid biosynthesis can be completely eliminated in C. tepidum shows that, in principle, the remaining unidentified carotenogenic enzymes can also be identified by gene inactivation.

MATERIALS AND METHODS

Organism and growth conditions. The strain of C. tepidum used was the plating strain WT2321 (58) derived from strain TLS (ATCC 49652) (57). The genome of the latter strain has been sequenced (15). Cells were grown in liquid CL medium or on solid CP medium as previously described (19, 25). Cells in liquid medium or on solid medium were incubated at 42 °C under incandescent illumination (approximately 150 μmol of photons m⁻² s⁻¹) unless stated otherwise. Cells for pigment analysis were grown in 25-ml cultures to the early stationary phase before analysis. Growth rates were measured in 25-ml cultures at 25 °C under vacuum and redissolved in a small volume of acetone-methanol (7:2, vol/vol) before analysis. In either case, the pigment extract was supplemented with 0.1 volume of 1 M ammonium acetate prior to injection to improve peak resolution. The HPLC column was a Discovery C₁₈ column (25 cm by 4.6 mm; 5-μm-diameter beads) (Supelco, Bellefonte, Pa.). The flow rate was 1 ml min⁻¹, and the mobile phase was composed of solvent A (methanol-acetonitrile-water [42:33:25 by volume]) and solvent B (methanol-acetonitrile-ethyl acetate [50:20:30 by volume]) as follows: 30% solvent B at the time of injection, linear increase to 100% solvent B in 52 min, constant at 100% solvent B for 6 min, and return to 30% solvent B in 2 min. The HPLC-DAD system consisted of an Agilent 1100 series binary pump (model G1312A), vacuum degasser (model G1379A), manual injector (model G1328A), and diode-array detector (model G1315B), and the data acquired were analyzed with ChemStation for LC 3D software (version A.08.03; Agilent Technologies, Waldbronn, Germany). The HPLC-MS system consisted of a Micromass Quattro II mass spectrometer equipped with a Shi-madzu LC10ADVP pump (Shimadzu, Columbia, Md.). Samples were analyzed using atmospheric pressure chemical ionization in negative ion mode with a corona needle potential of −4.0 kV, and the data acquired were analyzed with MassLynx software (version 3.5; Micromass, Ltd., Manchester, United Kingdom).

Pigments were identified by a combination of elution time, absorption spectrum, and molecular mass analysis and by comparison to the previously determined carotenoid composition of wild-type C. tepidum (53). The chromophores used for identification (all-trans isomers) and absorption maxima were as follows: chlorobactene and γ-carotene, 436 (shoulder), 461, and 491 nm; lycopene, 447, 472, and 503 nm; neurosporenne, 417, 440, and 469 nm; γ-carotene, 380, 401, and 425 nm; phytoene, 276 (shoulder), 287, and 298 nm; BChl a, 435 and 667 nm; and BChl a, 364, 602, and 770 nm. Absorption coefficients (in liters per gram per centimeter) used for quantification (22, 51) follow: for chlorobactene and γ-carotene, 265 at a wavelength of 490 nm; for glucoside derivatives of OH-chlorobactene and OH-γ-carotene, 198 at 490 nm; for glucoside laurate ester derivatives of OH-chlorobactene and OH-γ-carotene, 158 at 490 nm; for lycopene, 345 at 470 nm; for γ-carotene, 256 at 400 nm; for phytoene, 125 at 287 nm; for BChl c, 86 at 667 nm; and for BChl a, 60 at 770 nm. Retention times and masses of individual pigments are given below (see Table 1).

Sequence analysis and phylogenetic trees. Protein sequences were aligned and analyzed with ClustalX (version 1.83) (http://www.ibgc.u-strasbg.fr/BioInfo /ClustalX/) (55), MacVector (version 7.0; Genetics Computer Group), and TreeView (version 1.6.6) (http://taxonomy.zoology.gla.ac.uk/rod/treview.html) (45). Unrooted phylogenetic trees for presentation were constructed with PAUP (version 4.0; Sinauer Associates, Sunderland, Mass.).

RESULTS

Identification of genes involved in carotenoid biosynthesis.

On the basis of sequence similarity with enzymes known to be involved in carotenoid biosynthesis in other organisms, genes encoding the following proteins were tentatively identified in the C. tepidum genome sequence (15): one CrtB-type phytoene synthase (CT1386), two CrtP/CrtQ-type plant or cyanobacterial phytoene desaturases (CT0807 and CT1414), two proteins related to CrtH-type plant or cyanobacterial carotenoid cis-trans isomerases (CT0180 and CT0649), one CrtC-type purple bacterial carotenoid hydratase (CT0301), and one CrtU-type β-carotene desaturase (CT0323). Except for crtC/CT0301, which is located immediately upstream of the petC and petB genes that encode the Rieske iron-sulfur protein and the associated cytochrome b, none of the genes encoding these proteins are clustered with any other genes obviously related to pho-
tosynthesis. A putative glycosyl transferase (CT1416) was identified immediately upstream of chlorosome protein CsmH (CT1417) and downstream of crtQ/CT1414; because of this gene organization, CT1416 was hypothesized to encode an OH-carotenoid glucosyltransferase. No homolog of any known lycopene cyclase was found. A putative flavoprotein (CT1357) with little similarity to any proteins in the databases was hypothesized to be a possible lycopene cyclase or carotenoid...

FIG. 1. Proposed pathway of carotenoid biosynthesis in C. tepidum. The HPLC peak numbers are shown in parentheses after the chemical names. All genes shown in this scheme have been inactivated. Geranylgeranyl-PP, geranylgeranyl diphosphate.

FIG. 2. Maps illustrating the DNA constructs used to inactivate genes. The small labeled arrows indicate the binding sites of the primers used to generate PCR products. The gene-inactivating constructs were made by fusion of the PCR products with a fragment containing the aadA antibiotic resistance marker by megaprimer PCR or by ligation. See text and reference 25 for details.
TABLE 1. Pigment composition of wild-type and mutant strains of C. tepidum

<table>
<thead>
<tr>
<th>Pigment (peak no.)</th>
<th>Retention time (min)*</th>
<th>Mass (m/z)</th>
<th>Wild type</th>
<th>CT1386 (crtB)</th>
<th>CT1315 (crtP)</th>
<th>CT1414 (crtQ)</th>
<th>CT0323 (crtR)</th>
<th>CT0301 (crtU)</th>
<th>CT0649 (crtH)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phytoene (1)</td>
<td>54.7, 55.1</td>
<td>544.5</td>
<td>1</td>
<td>0</td>
<td>81</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>ζ-Carotene (2)</td>
<td>52.5, 52.9</td>
<td>540.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>48</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Lycopene (3)</td>
<td>49.9, 50.2</td>
<td>536.4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>21</td>
</tr>
<tr>
<td>ζ-Carotene (4)</td>
<td>51.6, 51.7</td>
<td>536.4</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>41</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Chlorobactene (5)</td>
<td>48.0, 48.2</td>
<td>532.4</td>
<td>47</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>45</td>
<td>4</td>
</tr>
<tr>
<td>1’, 2’-Dihydro-γ-carotene (6)</td>
<td>53.6, 53.8</td>
<td>538.5</td>
<td>tr</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>tr</td>
<td>tr</td>
</tr>
<tr>
<td>1’, 2’-Dihydrochlorobactene (7)</td>
<td>50.2, 50.4</td>
<td>532.4</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>tr</td>
</tr>
<tr>
<td>OH-γ-Carotene (8)</td>
<td>41.4</td>
<td>554.4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>OH-Chlorobactene (9)</td>
<td>36.6</td>
<td>550.4</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>OH-γ-Carotene glucoside laurate (10)</td>
<td>50.9</td>
<td>898.7</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>OH-Chlorobactene glucoside laurate (11)</td>
<td>47.6</td>
<td>894.6</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Total carotenoid* 58 0 81 65 55 52 39
BChl a (12) 34.9 910.5 19 11 13 13 18 16 17

* Retention times are from Fig. 4 and 5. More than one retention time is due to the presence of both cis and trans isomers.

* Each value represents the average of at least two measurements on the same cell culture (for values higher than 4, the standard deviation did not exceed 12% of the average); tr, trace.

* The total carotenoid amounts are shown in boldface type for emphasis. The amount of total carotenoid may be higher than the sum of the amounts of the listed carotenoids because not all carotenoid species are listed. See text for details.

1’,2’-saturate. All of the putative desaturases and isomerases (CT0180, CT0323, CT0649, CT0807, CT1357, and CT1414) contain a conserved, N-terminal flavin-binding motif: (A/V/L/I)(V/I)(L/F)X(G/A)G(V/L/I)(G/A)G(L/I)(A/S/T)X4LX8(V/L). This motif is characteristic of all known carotenoid desaturases and cis-trans isomerases (36, 50).

The incomplete genome sequence of Chloroflexus aurantiacus was also surveyed using data available from the websites of the Joint Genome Institute (Walnut Creek, Calif.) (http://www.jgi.doe.gov) and the National Center for Biotechnology Information (Bethesda, Md.) (http://www.ncbi.nlm.nih.gov). Full-length homologs of one CrtB-type phytoene synthase, one CrtI-type phytoene desaturase (ORF2 in Fig. 3C), one CrtO-type carotenoid ketolase (ORF1 in Table S3 in the supplemental material), and one CrtY-type lycopene cyclase were found. In addition, truncated homologs of one CrtU-type carotenoid desaturase (ORF3 in Fig. 3C) and one CrtP-type carotenoid desaturase (ORF4 in Table S3 in the supplemental material) were also found. (For detailed sequence information and accession numbers, see Table S3 in the supplemental material.)

Phylogenetic analysis of carotenogenic enzymes. Figure 3 shows unrooted phylogenetic trees for most of the carotenogenic enzymes identified in C. tepidum and Chloroflexus aurantiacus. The trees show that the sequences from C. tepidum (including that of CrtC; data not shown) are not closely related to the sequences from Chloroflexus aurantiacus or any other organism. A CrtU desaturase produces aromatic carotenoids in the actinomycetes Mycobacterium aurum (56), Streptomyces griseus (36), and Brevibacterium linens (35) and in C. tepidum (this work; see below). Surprisingly, CrtU homologs were also identified in four cyanobacteria (Fig. 3B). Cyanobacteria do not produce aromatic carotenoids, and the function of the CrtU homolog in these bacteria is not known. Interestingly, the CrtU sequences from C. tepidum and the cyanobacteria contain a Rieske iron-sulfur domain of 116 amino acids in the middle of the sequence; this domain is not found in the CrtU sequences from actinomycetes. This suggests that the CrtU enzymes in C. tepidum and cyanobacteria may have a common evolutionary origin. The CT0180 protein in C. tepidum has one homolog with unknown function in each of the eight cyanobacteria whose genomes have been completely sequenced (Fig. 3C) (based on information available from the Cyanobase database [Kazusa DNA Research Institute, Kisarazu, Japan] [http://www.kazusa.or.jp/cyano]); however, no close homologs occur in any other organism.

Pigment analysis of the wild type and crt mutants of C. tepidum. The levels of BChl c in cells of the wild type and all mutants were very similar as judged from the absorption spectrum of intact cells (data not shown). In addition, no significant changes were observed in the absorption properties of BChl c in intact cells of the mutants. Thus, the absence of carotenoids did not have an effect on BChl c biosynthesis or on the aggre-
toene desaturase. Because the major carotenoid in the CT1414 mutant was \(/\text{H}9256\)-carotene, it was logical to identify CT0807 as \(\text{crtP}\), encoding a phytoene desaturase that produces \(/\text{H}9256\)-carotene and to identify CT1414 as \(\text{crtQ}\), encoding a \(/\text{H}9256\)-carotene desaturase that produces lycopene. The CT1414 mutant also accumulated small amounts of phytoene and a compound that was tentatively identified as 1,2-dihydro-\(/\text{H}9256\)-carotene.

The dominant carotenoid in a cell extract of the CT0649 mutant was all-trans-lycopene (eluting at 50.2 min; absorption maxima of 474, 472, and 503 nm; Fig. 4). Another component eluting at 49.9 min had the same mass as lycopene and absorption maxima of 445, 467, and 499 nm and was tentatively identified as a cis isomer of lycopene. Small amounts of phytoene, chlorobactene, \(/\text{H}9253\)-carotene, and 1,2-dihydrolycopene were also found. Evidence of OH-lycopene or derivatives of OH-lycopene was not found. These observations strongly suggest that CT0649 encodes a CrtH-type lycopene cis-trans isomerase if one considers the following. Desaturation of \(/\text{H}9256\)-carotene by CrtQ produces a cis-lycopene and most likely takes place in the cytoplasmic membrane (7, 12). A mutant lacking \(\text{crtH}\) may therefore be expected to accumulate cis-lycopene, because all known lycopene cyclases act only on all-trans-lycopene (30). Although cis-lycopene is relatively stable by itself, it is readily isomerized to all-trans-lycopene when exposed to light and a photosensitizing molecule like BCHls (30). However, most of the cis-lycopene produced in the CT0649 mutant is probably translocated to the chlorosome without being isomerized prior to translocation. Once in the chlorosome, this cis-lycopene is photoisomerized to all-trans-lycopene due to photosensitization by BCHl \(c\) but is then unavailable for further enzymatic modification. Although it was not explicitly shown that lycopene in the CT0649 mutant accumulated in the chlorosomes, it was shown that phytoene in the CT0807 mutant predominantly accumulated in the chlorosomes (data not shown), thus suggesting that most carotenoids in all \(\text{crt}\) mutants are correctly translocated to the chlorosomes. A small proportion of the cis-lycopene produced in the CT0649 mutant is probably photoisomerized in the cytoplasmic membrane before being translocated to the chlorosome and thus would be available to lycopene cyclase and other enzymes. This probably accounts for the small amounts of chlorobactene and \(/\text{H}9253\)-carotene found in the CT0649 mutant.

The CT0301 mutant lacked OH-chlorobactene and OH-\(/\text{H}9253\)-carotene and their derivatives; therefore, the CT0301 gene was identified as \(\text{crtC}\), which encodes a carotenoid 1’,2’-hydrolase. The CT0323 mutant lacked chlorobactene and all its derivatives and accumulated only \(\gamma\)-carotene and its derivatives. The CT0323 gene was thus identified as \(\text{crtU}\) encoding \(\gamma\)-carotene desaturase. No changes in the carotenoid composition of the CT0180, CT1357, and CT1416 mutants were detected, which
means that these three genes probably do not function in carotenoid biosynthesis.

Upon careful analysis of the HPLC traces, an unidentified component named F (for flexirubin-like) with an elution time of 49.3 min and absorption peaks at 285, 325, and 453 nm was found (Fig. 6). This component F was found in the CT1386, CT0807, CT1414, CT0323, and CT0649 mutants but was not unambiguously identified in the CT0301 and CT0180 mutants or in the wild type due to masking by carotenoids. However, previous analyses using different HPLC systems suggest that component F is also present in the wild type (N.-U. Frigaard,
unpublished data). The absorption band around 453 nm resembles that of a conjugated system of nine double bonds, and due to the lack of fine structure, the spectral data suggest the presence of at least one carbonyl group (51). Component F is not likely to be a carotenoid, because it was present in the presence of at least one carbonyl group (51). Component F resembles those of the flexirubin-type pigments found in some Cytophaga, Flavobacterium, and Flexibacter species (1, 51), to which C. tepidum is distantly related on the basis of 16S rRNA and whole-genome phylogeny (15). Flexirubin-type pigments are not derived from isoprenoid precursors and consist of a polyenoic acid chromophore terminated by a hydroxyphenyl group and esterified with an alkyl-substituted phenol group of the resorcinol type. The function of these pigments is not known. If component F has a specific absorption coefficient of about 200 liters g⁻¹ cm⁻¹ at a wavelength of 453 nm, cell extracts of C. tepidum contain roughly 1 mg of component F per g of BChl c. Analyses show that component F is present in isolated chlorosomes (data not shown).

Inhibition of lycopene cyclization in C. tepidum. The gene encoding lycopene cyclase in C. tepidum was not identified in this study. Nevertheless, it was observed that the addition of the herbicide 2-(4-methylphenoxy)triethylamine hydrochloride (MPTA) to the growth medium of C. tepidum inhibited lycopene cyclization (data not shown). MPTA has been shown to inhibit lycopene cyclization in both microbes and plants (7). Lycopene constituted about 80% of the total carotenoids in cultures of wild-type C. tepidum containing 3 μM MPTA and about 90% of the carotenoids in cultures containing 100 μM MPTA (the remaining carotenoid mostly being chlorobactene). The growth rate and coloration of wild-type C. tepidum cultures were not significantly affected by MPTA concentrations between 3 and 100 μM. MPTA also inhibited lycopene cyclization in a bchK mutant of C. tepidum. This mutant does not contain BChl c but contains various carotenoids in amounts similar to those in the wild type (22). When the bchK mutant was grown in the presence of 3 μM MPTA, about 85% of the total carotenoid pool was lycopene. In contrast to the wild type, the bchK mutant grew more slowly when the MPTA concentration was 3 μM or higher and grew very little when the MPTA concentration was 100 μM. Nicotine has also been shown to inhibit lycopene cyclization in C. tepidum (53). Isolation of C. tepidum mutants resistant to these chemical inhibitors of lycopene cyclization may help in identifying the genetic basis of lycopene cyclization.

Effects of crt mutations on the growth of C. tepidum. No effects on cell morphology in the crt mutants were observed by light microscopy (data not shown). This contrasts with observations of Chlorobium phaeobacteroides, which forms filaments when cells are treated with the carotenoid biosynthesis inhibitor 2-hydroxybiphenyl (4, 5). Thus, filament formation is probably due to secondary effects of the inhibitor on cell morphology and division. Chlorosomes are apparently formed normally in the C. tepidum mutants, since chlorosomes similar to wild-type chlorosomes were isolated from the crtB and crtP mutants (data not shown). Chlorosomes from the crtP mutant contained most of the phytolene present in the cells, demonstrating that this carotenoid precursor is translocated to the chlorosomes, as are most of the carotenoids in the wild type.

Figure 7 shows the growth rates of the wild type and various crt mutants of C. tepidum at three different light intensities. The growth rates of four mutants (crtU, crtH, CT0180, and CT1416 mutants) showed no significant deviation from the wild-type growth rate at any light intensity (data for the CT0180 and CT1416 mutants not shown). The growth rate of the crtC mutant differed from the wild-type growth rate only at the lowest light intensity (10 μmol of photons m⁻² s⁻¹, limiting for growth in all strains), when it grew only about 60% as fast as the wild type. The growth rates of the three other mutants (crtB, crtP, and crtQ mutants) were significantly lower than the wild-type growth rate at all light intensities. At the lowest light intensity, the growth rates of the crtP and crtQ mutants were about 40% that of the wild type, and the growth rate of the crtB mutant was only about 25% that of the wild type. At increasing light intensities, the growth rates of the crtB, crtP, and crtQ mutants were higher but did not reach the wild-type growth rate, even at the highest light intensity examined (588 μmol of photons m⁻² s⁻¹), when these mutants grew about 65% as fast as the wild type.

DISCUSSION

Carotenoid biosynthesis in green sulfur bacteria. On the basis of the mutational studies in C. tepidum in this work, we propose the carotenoid biosynthetic pathway shown in Fig. 1. The intermediates and enzymes leading to lycopene are identical to those in cyanobacteria and plants (7, 12). A CrtU-like enzyme carries out the desaturation of 3-carotene to chlorobactene; this enzyme is related to but distinct from the enzyme found in actinomycetes. A CrtC enzyme, similar to that in purple bacteria, hydrates both 3-carotene and chlorobactene. Since hydrated species of lycopene or earlier intermediates in the pathway were not found in the wild type or any mutant,
CrtC apparently cannot efficiently hydrate these species. There appears to be little or no feedback inhibition or obligatory multienzyme complex formation in this pathway, since the genetic inactivation of one enzyme did not significantly affect the apparent activity of the remaining enzymes or the total carotenoid content (Table 1).

Four enzymes are still unidentified in the proposed pathway in Fig. 1. These unidentified enzymes include the glycosyltransferase and a laurate transferase that are required to produce the OH-carotenoid glucoside laurate esters (53). It is noteworthy that the enzymes that form the glucoside laurate esters seem to have higher affinity for γ-carotene than chlorobactene, since the ratio of chlorobactene to γ-carotene ranges from 10 to 20, whereas the ratio of OH-chlorobactene glucoside laurate to OH-γ-carotene glucoside laurate ranges from 1 to 2.

Another enzymatic activity not yet completely understood is the carotenoid 1,2-saturase, which produces 1',2'-dihydrochlorobactene and 1',2'-dihydro-γ-carotene (53). In *C. tepidum*, 1',2'-dihydrochlorobactene can be a major carotenoid species under certain growth conditions, such as prolonged incubation (N.-U. Frigaard, unpublished data) or low light intensities (10). The only other organism in which 1,2-dihydrocarotenoids have been detected is *Blastochloris* (formerly *Rhodopseudomonas*) *viridis* (41). The major carotenoid in this organism is 1',2'-dihydroxanthophyll, and 1,2-dihydrocarotenoids may constitute up to 94% of the total carotenoids. Insufficient DNA sequences are currently available from the photosynthesis gene cluster of *Blastochloris viridis* to allow a detailed comparison with the *C. tepidum* sequences.

Two paralogous genes (*bchP/CT2256* and *bchO/CT1232*) in *C. tepidum* encode putative geranylgeraniol reductases involved in the BchP enzyme involved in BChl *a* biosynthesis in purple bacteria (15, 23). The *bchP* and *bchO* genes have been inser- tionally inactivated in *C. tepidum* to produce two independent mutants (A. Gomez Maqueo Chew, N.-U. Frigaard, and D. A. Bryant, unpublished data). Both the *bchP* and *bchO* mutants produce altered BChl *a* and Chl *a* species that are consistent with the products of these genes functioning in reduction of the geranylgeranyl tail of (B)Chls. Additionally, the *bchP* and *bchO* mutants are deficient in 1',2'-dihydrochlorobactene. Thus, it appears that BchP and BchO, in some as yet poorly understood but interdependent manner, are responsible for the saturation of the isoprenoid moiety of both BChl *a* and Chl *a*, and carotenoids in *C. tepidum*. Since it is not simple to dis-tinguish between the phenotypic effects of modified (B)Chls and modified carotenoids in the *bchP* and *bchO* mutants, these mutants were not further characterized in this work. Further studies with these mutants to resolve these issues are in progress.

The only other important carotenogenic enzyme not yet identified in *C. tepidum* is lycopene cyclase. The genome does not encode any homolog of the three known types of lycopene cyclases (50). Interestingly, some cyanobacteria also lack an identifiable lycopene cyclase. Given the similarity of many of the carotenogenic enzymes and other proteins involved in photosynthesis between *C. tepidum* and cyanobacteria (23), the possibility that *C. tepidum* and some cyanobacteria share a novel type of lycopene cyclase is intriguing. Since four cyanobacteria in Fig. 3B (*Synechocystis* sp. strain PCC 6803, *Nostoc* sp. strain PCC 7120, *Gloeobacter violaceus,* and *Trichodesmium erythraeum*) also lack an identifiable lycopene cyclase, the possibility that the CrtU homologs of these cyanobacteria could function as lycopene cyclases is interesting. Consistent with this suggestion, other cyanobacteria that have an identifiable lycopene cyclase (for example, three *Prochlorococcus marinus* strains and *Synechococcus* sp. strain WH8102) lack homologs of CrtU (based on information available from the CyanoBase database).

**Carotenoid biosynthesis in green filamentous bacteria.** The only organisms other than green sulfur bacteria that contain chlorosomes and BChl *e* are the green filamentous bacteria, represented by *Chloroflexus aurantiacus* (28, 32). However, the biosynthetic pathway leading to lycopene and its cyclization products in *Chloroflexus aurantiacus* appears to be different from the pathway in *C. tepidum*. The portion of the *Chloroflexus aurantiacus* genome that has been sequenced contains one or possibly two genes encoding homologs of CrtI-type carotenoid desaturases and possibly one gene encoding a homolog of a CrtP-type desaturase, but it does not encode a homolog of the carotenoid cis-trans isomerase. *Chloroflexus aurantiacus* may thus employ a classical bacterial CrtI-dependent phytoene desaturase (see introduction). Alternatively, *Chloroflexus aurantiacus* may employ a hybrid pathway like *Nostoc* sp. strain PCC 7120, which uses a CrtP-like phytoene desaturase that produces ψ-carotene and a ψ-carotene desaturation-related to CrtI proteins (sometimes denoted CrtQ or CrtQa [Fig. 3C]) (2, 40, 50). The major carotenoids in *Chloroflexus aurantiacus*, β-carotene and γ-carotene, are apparently formed by a classical bacterial lycopene cyclase of the CrtY family. Under oxic conditions, carotenoids that contain 4-oxo-β end groups, including echinenone (4-oxo-β-carotene) and myxobactene (4-oxo-γ-carotene glucoside) (14, 31, 47), are probably formed by a CrtQ ketolase similar to those in cyanobacteria (16). (The GenBank accession numbers of all the *Chloroflexus aurantiacus* protein sequences mentioned in this paragraph are listed in Table S3 in the supplemental material.)

**Effect of carotenoid composition on growth of *C. tepidum***

The growth rates of two mutants with altered carotenoid composition (*crtU* and *crtH* mutants) were not significantly different from those of the wild type at any light intensity (Fig. 7). Thus, the absence of aromatic carotenoids (chlorobactene and its derivatives) in the *crtU* mutant was not a significant disadvantage under the growth conditions tested. Likewise, elimination of most of the cyclic carotenoids and the accumulation of lycopene in the *crtH* mutant did not significantly affect the growth of this strain. The absence of hydroxylated carotenoids and their derivatives (glucosides and fatty acid glucoside esters) in the *crtC* mutant appeared to cause a growth deficiency only at limiting light intensity (10 μmol of photons m⁻² s⁻¹). This could be due to reduced carotenoid level or efficiency of the reaction centers, because most of the hydroxylated carotenoids and their derivatives are associated with the reaction centers (54). In contrast, growth was impaired in the *crtB*, *crtP*, and *crtQ* mutants under all light intensities examined (Fig. 7). These three mutants share the common property that they lack colored carotenoids, i.e., carotenoids with a chromophore consisting of nine or more conjugated double bonds. The wild type and all *crt* mutants with no growth rate phenotype have colored carotenoids (see chemical structures in Fig. 1).
It is not yet clear what causes the decreased growth rates of the \( \text{crtB} \), \( \text{crtP} \), and \( \text{crtQ} \) mutants. It is also not clear whether the cause of the decreased growth rate under low light intensities is the same as the cause for the decreased growth rate under high light intensities. However, it is striking that the three mutants with decreased growth rates (\( \text{crtB} \), \( \text{crtP} \), and \( \text{crtQ} \) mutants) also have decreased BChl \( a \) levels (Table 1). Although BChl \( a \) is not the major light-harvesting pigment, it is essential for the transfer of excitation energy from the major light-harvesting pigment (BChl \( c \)) in the chlorosome to the reaction center. Thus, the decreased growth rates could be due to less efficient energy conservation because of decreased BChl \( a \) levels. How the absence of certain carotenoids could cause a decrease in the BChl \( a \) level is not known, but there may be structural or functional requirements for such carotenoids in the BChl \( a \)-containing CsmA protein in the chlorosome baseplate (13, 43, 49) or in the reaction center (54) or both. Both spectroscopic analyses (3, 6, 42) and chemical analyses (13, 43) suggest that BChl \( a \) and carotenoids are intimately associated in the chlorosome baseplate and that the CsmA protein probably binds to at least one carotenoid molecule. Other analyses also suggest a correlation between colored carotenoids, BChl \( a \), and CsmA. Chlorosomes isolated from \( \text{C. phaeobacteroides} \) and \( \text{Chloroflexus aurantiacus} \) grown in the presence of the phytone desaturase inhibitor 2-hydroxybiphenyl are almost completely devoid of colored carotenoids, and the levels of BChl \( a \) and CsmA in such cells are decreased by roughly 35 to 60% (4, 5, 17). Analogously, when \( \text{crtI} \) is inactivated in the purple bacterium \( \text{Rhodobacter sphaeroides} \) (which then accumulates colorless phytone), formation of light-harvesting complex 2 is blocked, whereas formation of light-harvesting complex 1 is not (11, 37, 46).

Another possibility is that excited triplet states of BChls accumulate in the \( \text{crtB} \), \( \text{crtP} \), and \( \text{crtQ} \) mutants and that these BChl triplets cause damage to the photosynthetic apparatus. Quenching of BChl triplets requires a carotenoid with a minimum of nine conjugated double bonds (18). Such carotenoids are absent only in the \( \text{crtB} \), \( \text{crtP} \), and \( \text{crtQ} \) mutants, and these mutants should thus accumulate increased levels of BChl \( c \) and BChl \( a \) excited triplet states. Both BChl \( c \) and BChl \( a \) triplets have been detected in both cells and isolated chlorosomes of green sulfur bacteria (3, 6, 42). BChl triplets are known to be detrimental for photosynthetic organisms under aerobic conditions because of singlet O\(_2\) formation (18). However, this cannot explain the growth deficiency of the \( \text{crtB} \), \( \text{crtP} \), and \( \text{crtQ} \) mutants, because they grow only under strictly anaerobic conditions. Nevertheless, accumulation of high levels of BChl \( c \) and BChl \( a \) triplets in \( \text{C. tepidum} \) may damage the cells in a yet unidentified manner.

Although carotenoids may certainly play a role in light harvesting in natural environments, it is unlikely that the growth defects observed here at low light intensity are due to reduced light harvesting in the mutants lacking colored carotenoids. The tungsten lamps used to provide the illumination in the growth experiments emit relatively little light with wavelengths shorter than 600 nm. In order to explore the roles of carotenoids in light harvesting, future studies will systematically study the growth behavior of the mutants under different light regimes.

**Evolution of carotenogenic enzymes.** It is not clear why two different biosynthetic pathways for lycopene exist in nature nor is it clear what advantages each may present over the other. The lycopene biosynthetic enzymes (CrtP, CrtQ, and CrtH) in green sulfur bacteria and cyanobacteria are obviously related, although very distantly (Fig. 3C and D), and they have thus far been found only in these two bacterial groups (photosynthetic eukaryotes aside). It is therefore likely that these enzymes originated in a common ancestor of modern-day green sulfur bacteria and cyanobacteria and were inherited vertically. This common ancestor probably contained an ancestral form of a type I photosynthetic reaction center (33), since the CrtP, CrtQ, and CrtH enzymes occur only in organisms containing type I reaction centers.

* cis* carotenoids appear to perform important functions in reaction centers (34). It has been suggested that all reaction centers contain 15-*cis* carotenoid isomers, which can be generated by thermal isomerization of the corresponding all-*trans* isomers (34). In addition, photosystem I from cyanobacteria contains 9-*cis* carotenoids, which are not found in type II reaction centers (27, 34). The presence or absence of 9-*cis* carotenoids in green sulfur bacterial reaction centers has not been conclusively demonstrated. Since 9-*cis* carotenoids are not easily formed by thermal isomerization, the enzymatic formation of *cis* carotenoids other than the 15-*cis* isomers may have provided a selective advantage to the ancestral type I reaction center. Because the CrtI-dependent biosynthetic pathway for lycopene does not produce *cis*-isomer intermediates, we propose that the evolution of the CrtP/CrtQ/CrtH-dependent pathway is linked to the postulated advantage of having particular *cis* carotenoids in the ancestral type I reaction centers.

On the basis of sequence similarity, CrtH and CrtI are distantly related (Fig. 3C), and CrtH probably evolved from the much more widespread CrtI (50). Analogously, on the basis of sequence similarity, CrtP, CrtQ, and CrtU are distantly related (50). CrtP and CrtQ probably arose by an ancient gene duplication event. Sandmann has proposed that CrtU originated in the actinomyces and that CrtP and CrtQ evolved from CrtU after horizontal transfer to cyanobacteria (50). However, since the function of CrtU in green sulfur bacteria and actinomycetes is the same but that CrtU proteins from green sulfur bacteria and cyanobacteria are more similar structurally, it seems unlikely that CrtU has been exchanged horizontally between actinomycetes and cyanobacteria. We therefore propose that the divergence of CrtP/CrtQ and CrtU occurred in an ancestor of the green sulfur bacteria.

**ACKNOWLEDGMENTS**

This work was supported in part by grant DE-FG02-94ER20137 to D.A.B. from the U.S. Department of Energy.

We thank Francis X. (Buddy) Cunningham, Jr. (University of Maryland), for supplying the herbicide MPTA, Kazuhiro Inoue (Kanagawa University, Kanagawa, Japan) for sharing unpublished DNA sequences from *Blastochloris viridis*, and Harry A. Frank (University of Connecticut) for discussions on carotenoids.

**REFERENCES**


APPENDIX C

ATTEMPTS TO TRANSFORM *CHLOROBIUM PHAEOBACTEROIDES* STRAIN 1549 WITH EXOGENOUS DNA
Abstract

In order to comprehensively investigate the genes involved in bacteriochlorophyll (BChl) e and isorenieratene biosynthesis, it would be helpful to be able to inactivate genes in a BChl e-producing strain of green sulfur bacteria (GSB). The strain *Chlorobium (Chl.) phaeobacteroides* strain 1549 produces BChl e as its primary antenna BChl and, like other transformable GSB strains, this organism utilizes thiosulfate as an electron donor and grows well on solid medium. Attempts to introduce foreign DNA into these cells via natural transformation, electroporation, and conjugation are described. At the current time, none of these methods has successfully resulted in inactivation of the *bchU* gene used as test case.
C.1 Introduction

In the studies of pigment biosynthesis in green sulfur bacteria (GSB), the existence of a genetic system for targeted inactivation of non-essential genes has been invaluable (3-7, 10-12). However, all strains in which genes can presently be inactivated are green-colored (2, 9, 14). Thus, analysis of genes required for biosynthesis of the pigments produced only by brown-colored species of GSB, isorenieratene and BChl e, can at the moment only be performed by heterologous expression of those genes in E. coli or Chlorobium (Chl.) tepidum, a green-colored, BChl c-producing species (see chapters 3 and 6) or by biochemical inhibition of specific protein activities (see chapter 3).

The two GSB species in which genes can be targeted for inactivation, Chl. tepidum strain TLS and Chl. vibrioforme strain 8327, have unique metabolic characteristics that make selection of transformants possible. They both grow well on solid medium, which is possible because they can utilize thiosulfate, rather than sulfide, as an electron donor. In these two species, foreign DNA has been introduced using natural transformation (1, 2), electroporation (9), and conjugation (14).

Chl. tepidum has been the most frequently studied organism among GSB for genetic studies. Three antibiotic resistance markers have been successfully used in this species (2): aadA derived from the plasmid pSRA2 or pSRA81, conferring resistance to streptomycin and spectinomycin; aacC1 derived from the plasmid pMS255 and conferring resistance to gentamicin; and erm derived from the plasmid pRL161 and conferring resistance to erythromycin. The cat cassette also derived from plasmid pRL161 and conferring resistance to chloramphenicol has not been successful used; although a construct (pCMR325) exists in which the cat gene is under the control of the promoter found in the aadA cassette, this construct has not yet been tested in Chl. tepidum.

When the bchU gene, encoding the BChl c- and e-specific C20 methyltransferase, is inactivated in Chl. tepidum, the resulting mutant strain produces BChl d (12; see chapter 5). When grown at saturating light intensities, the BChl d-producing mutant is not impaired in growth compared to wild type, and at low light intensities, although it grows more slowly, it is still viable. Therefore, the bchU gene is not required for growth of Chl. tepidum. The hypothetical pigment BChl f, which would be unmethylated at the C20
position and formylated at the C7 position, has been synthesized and analyzed in vitro (13) but has been never been observed in nature. The bchU gene has been amplified by PCR and sequenced from Chl. phaeobacteroides strain 1549 (12, NCBI accession # AY452767; see chapter 5), where it presumably has the same function as in Chl. tepidum and Chl. vibrioforme strain 8327c. This gene was chosen as the test platform for gene inactivation in Chl. phaeobacteroides strain 1549. The absence of BchU should not affect photosynthesis or growth under normal growth conditions, as in both Chl. tepidum and Chl. vibrioforme. Additionally, construction of a GSB strain able to synthesize BChl f in vivo would make physiological experiments with this strain possible, and perhaps these studies would provide an explanation for the absence of this pigment in natural environments.

C.2 Methods
C.2.1 Strains and growth conditions. The strain used was Chlorobium phaeobacteroides strain 1549, which was isolated from Lake Kinneret, Israel (John G. Ormerod, personal communication). Manipulations of this strain were performed in an anaerobic chamber (Coy, Grass Lakes, MI) with an atmosphere of 10% H₂, 10% CO₂, and 80% N₂ (v/v). Cells were grown either in liquid CL or on solid CP as for Chl. tepidum at temperatures of 28-30°C.

When cells of a dense culture of Chl. phaeobacteroides strain 1549 are centrifuged, the pellet is loose and many cells remain in the supernatant. It was discovered that cells in early exponential phase form more tightly packed pellets and fewer of them are lost. For all of the experiments described below, each was performed using both dense cultures (OD₆₆₀nm >2) of Chl. phaeobacteroides strain 1549 and early-exponential phase (OD₆₆₀nm ~0.3-0.5) cultures.

Routine recombinant DNA procedures were carried out in E. coli strains DH10B or DH5α, as appropriate. All cells were grown in liquid LB or solid LB-agar amended with 100 µg ampicillin ml⁻¹, 25 µg chloramphenicol ml⁻¹, 30 µg kanamycin ml⁻¹, or 50 µg streptomycin ml⁻¹ and 100 µg spectinomycin ml⁻¹, as appropriate. For conjugation from E. coli, the strain S17-1 was made electrocompetent, electroporated with the plasmid to be transferred, and grown in LB amended with the appropriate antibiotics.
C.2.2 Antibiotic resistance of *Chl. phaeobacteroides* strain 1549. CP plates were prepared with a range of concentrations of streptomycin and spectinomycin together, gentamicin, erythromycin, and chloramphenicol (see Table C.1). Cells were grown until mid-exponential phase without antibiotics in liquid CL, then diluted 1:10, 1:10^3, and 1:10^5 and 50 µl of each dilution was spread on the plates. Plates were incubated for 2 weeks or until colonies appeared, and growth was scored.

C.2.3 Constructs for inactivation of *bchU* by natural transformation or electroporation. The *bchU* gene was amplified from genomic DNA of *Chl. phaeobacteroides* strain 1549 using primers BCHU FE (CGG GGT GAA TTC GGA CAG GCT GGA TAA C) and BCHU RP (GCC GAG CAC CCT GCA GGG CAT TCC), with EcoRI and PstI sites, respectively (underlined). The PCR product and plasmid pUC19 were both digested with EcoRI and PstI and after gel-purification of the fragments, were ligated together. The plasmids pSRA2, pMS266, and pCMR325 were digested with KpnI and the fragments corresponding to the antibiotic resistance cassettes were excised. The *bchU* gene of *Chl. phaeobacteroides* has an internal KpnI site, into which the antibiotic resistance cassettes were inserted to produce plasmids pCPB-A(+), pCPB-A(-), pCPB-G(+), pCPB-G(-), pCPB-C(+) and pCPB-C(-) (Figure C.1). The (-) and (+) designations indicate whether the cassette is transcribed in the opposite direction of *bchU*, (-), or the same direction, (+). These plasmids were amplified in 1 L of *E. coli* strain DH10B and purified; a stock of ~1 ml of each, at concentrations of ~1 µg µl⁻¹, is stored in freezer F3 in the box labeled PLASMIDS.

C.2.4 Constructs for inactivation of *bchU* by conjugation. The six inactivation constructs described above were amplified by PCR from the above constructs using the M13 primers and inserted into the conjugation plasmid pKJS2 using the restriction sites EcoRI and PstI to produce the plasmids pKJ-CPB-A(+), pKJ-CPB-A(-), pKJ-CPB-G(+), pKJ-CPB-G(-), pKJ-CPB-C(+) and pKJ-CPB-C(-).

C.2.5 Natural transformation of *Chl. phaeobacteroides* strain 1549. The plasmids used for transformation were digested with the enzyme Scal and the linearized plasmids were gel-purified. Cells were grown either until dense or until early exponential phase. Cells (1 ml of the culture) was removed and centrifuged in the anaerobic chamber at 13000 rpm for ~3 min. The supernatant was removed and the cells were resuspended in 30-50 µl of
the transformation construct in sterile water. The resulting suspension was immediately spotted on to nonselective CP plates and the cells were allowed to grow for 24 or 72 h. After 24 or 72 h, the cells were transferred to selective plates using sterile loops and incubated for 2 weeks or until colonies appeared.

C.2.6 Constructs for inactivation of BE1. The plasmids pTA-BEDH (see Chapter 6) and pSRA2 were digested with Smal, which has recognition sites in both BE1 and the downstream dehydrogenase. The aadA fragment and the fragment corresponding to pTA-BEDH without the internal fragment were gel-purified as described above and ligated together to produce plasmid pTA-BEDH-A (Figure C.2).

C.2.7 Methylation of DNA for transformation. Methylation of DNA using *Chl. phaeobacteroides* strain 1549 cell lysates was performed using the method of Zhou (16). Briefly, cells were grown until mid-exponential phase, then centrifuged and washed in 0.5 volume cold buffer M. After centrifuging again, cells were resuspended in 0.5 volume cold buffer M and lysed by sonication. The lysate was centrifuged and the supernatant was used for the methylation reaction. Cell lysate (5 µl) and 0.2 mM S-adenosylmethionine was added per µg of DNA, and the reaction was incubated at 30°C for several hours. The DNA was then purified by phenol-chloroform extraction and ethanol precipitation, linearized with AhdI, and used in both natural transformation and electroporation of *Chl. phaeobacteroides* strain 1549.

Buffer M contained 50 mM Tris-Cl, pH 7.5, 50 mM NaCl, 10 mM EDTA, and 5 mM 2-mercaptoethanol.

C.2.8 Electroporation of *Chl. phaeobacteroides* strain 1549. Cells were grown until early exponential phase; aliquots (1ml) were removed and centrifuged anaerobically. To find the highest voltage that was minimally lethal, the cells were washed in 1 ml sterile, anaerobic water three times. After the third wash, the supernatant was removed to the degree possible, and sterile water (1 µl) was added to the tube. The cell/DNA suspensions were transferred inside the anaerobic chamber to electroporation cuvettes, which were sealed with Parafilm and removed from the anaerobic chamber for electroporation. The voltages tested were 0, 1.2, 1.5 and 2.0 kV. The cuvettes were then placed back into the anaerobic chamber and the cells were spread on nonselective plates and grown for 1 week. For transformation, the cells were prepared and washed as described above, but
linearized plasmid DNA (1 µl) was added to the loose cell pellet and mixed by stirring. Electroporation was carried out as described above with a voltage of 1.5 kV, and the cells were spotted on to nonselective plates and incubated for 24 h. After 24 h, the cells were transferred to selective plates as described in section C.2.5.

**C.2.9 Conjugation of Chl. phaeobacteroides strain 1549.** The protocol described in Wahlund and Madigan (14) was used, but cells were grown at ~30°C. Electropotentent S11 *E. coli* were transformed with pKJ-based plasmids listed above and grown with selection until dense, approximately 10 h. They were then diluted 1:10 in LB without antibiotics and grown at 37°C for 2 h. *Chl. phaeobacteroides* was grown until the OD$_{660}$nm was approximately 0.6. Each strain (1 ml) was washed once in sterile medium, then resuspended in CLY medium (1 ml) (CL without added sulfide and with 0.05% yeast extract) and mixed at a ratio of 3:2 *Chl. phaeobacteroides*: *E. coli*. The mixtures were then centrifuged, resuspended in 0.2 ml CLY, and the mixture (50 µl) was spotted onto nonselective CP plates amended with 0.05% (w/v) yeast extract. Ordinarily, GSB cells grown on solid medium are provided with sulfide by placing a tube of thioacetamide:1M HCl (1:10, w/v) inside the jar with the plates; however, the *Chl. phaeobacteroides*: *E. coli* mixture was grown without this, as the sulfide would poison the *E. coli*. After 24 h growth under these conditions, the cells were transferred to selective plates as described above and incubated for 2 weeks or until colonies appeared.

**C.2.10 PCR analysis of putative transformants.** When colonies appeared on the selective plates, they were streaked either on plates with the same concentration of antibiotics or, if the colonies appeared small or unhealthy, on plates with slightly reduced antibiotic concentrations. At the same time, cells from the same colonies were resuspended in sterile water. These were diluted 1:1 or 1:10 (depending on density of cells) in 0.1% (w/v) NP40 and boiled for ~1 min. to lyse the cells, which could be monitored by a change in color from brown to green. This crude lysate (1 µl) was used as template in PCR to check for insertion of the antibiotic resistance cassette into *bchU*. Primer pairs om-2s/om-2as (for *aadA*), GmR-F/GmR-R (for *aacC1*), or EmR-F/EmR-R (for *erm*) as appropriate were used to check for the presence of a specific gene encoding antibiotic resistance in the cells of a colony; the *bchU* locus was then amplified from the
templates that were positive for the antibiotic resistance cassette, using primers BCHU-FE and BCHU-RP.

C.2.11 Pigment analysis of putative transformants. For whole-cell absorption spectroscopy, cells were centrifuged and resuspended in *Chlorobium* harvest buffer (1 ml; 10 mM KH₂PO₄ and 50 mM NaCl, pH 7.0) and spectra from 350 to 800 nm were recorded on a Cary-14 spectrophotometer, updated for computerized data acquisition and analysis (On-Line Instrument Systems, Bogart, GA). For absorption spectroscopy of whole-cell extracts, pigments were extracted from cell pellets by sonication in acetone:methanol (7:2, v:v) and diluted 1:10 in methanol. Spectra were recorded as for whole-cell spectra. For HPLC analysis, the lysate prepared as above was filtered through a 0.2 μm filter and injected into the HPLC. The HPLC system and solvent gradients were exactly as described in Chapter 5.

C.3 Results

C.3.1 Selective antibiotic concentrations for Chl. phaeobacteroides strain 1549. The concentrations of antibiotics at which *Chl. phaeobacteroides* is sensitive are summarized in Table C.1. These concentrations, for the most part, are comparable to those that inhibit growth in *Chl. tepidum*.

C.3.2 Natural transformation. Natural transformation of *Chl. phaeobacteroides* strain 1549 was attempted several times using *bchU* as a platform, with the *aadA*, *aacC1*, and *cat* cassettes inserted into *bchU*. In addition, it was attempted once using *BE1* interrupted with *aadA*, both with and without methylation of the DNA before transformation. In all cases, growth on selective plates was strongest on CP amended with gentamicin, with some growth on CP amended with streptomycin and spectinomycin. Most of the growth on plates appeared to be from clumps of cells, rather than from single cells, as the colonies tended to be irregularly shaped and taller than they were wide. The results were the same whether or not the plasmid DNA used for the attempted transformation had been linearized.

PCR from the colonies that grew on the first passage on selective plates did not indicate that another gene had been inserted into the *bchU* locus or that the antibiotic resistance cassettes were present.
C.3.2 Electroporation. There was more growth on selective plates after electroporation than after natural transformation attempts. Additionally, PCR results indicated that at least the \textit{aacC1} cassette was present in the putative transformants (Figure C.3). However, PCR across the \textit{bchU} locus did not demonstrate any insertion (Figure C.3). Although the absorption spectra of whole-cell pigment extracts differed from wild-type cells (Figure C.4), when analyzed by HPLC it was clear that the BCHls produced by these strains did not differ from wild-type BCHls, and that the variation in the absorption spectra of whole cells can be attributed to the fact that cells that have been growing on plates for long periods of time frequently appear to have degraded BCHls and a different complement of carotenoids than freshly grown liquid cultures.

C.3.3 Conjugation. Colonies appeared on selective plates within 2 weeks of conjugation. PCR analysis of these colonies indicated that the \textit{aadA} cassette was present in 8 of the 15 streptomycin- and spectinomycin-resistant colonies tested, though \textit{aacC1} was not present in any of the gentamicin-resistant colonies. PCR across the \textit{bchU} locus showed only wild-type copies of \textit{bchU} (Figure C.5). To eliminate the possibility that some of the putative transconjugants were merodiploid mutants, and that PCR had selectively amplified the wild-type allele, the PCR was repeated using the primer pair \textit{bchU}-F1 and \textit{om2}-s, which would only produce an amplicon from mutant alleles. In fact, the results from this PCR analysis demonstrated that there were no mutant alleles in any of the colonies tested (Figure C.6). After one more passage on selective medium, the colonies checked by PCR no longer had \textit{aadA}; the earlier positive results were most likely due to DNA derived from \textit{E. coli} strain S17-1.

C.3.4 Transformations with methylated DNA. In order to test whether the obstacle to transformation might be an endogenous restriction-modification system, both natural transformation and electroporation trials were performed with the construct for inactivation of the gene possibly involved in BCHl \textit{e} biosynthesis, \textit{BE1}, and the dehydrogenase immediately downstream of it (see Chapter 6). Parallel experiments were performed in which this construct was methylated by a \textit{Chl. phaeobacteroides} strain 1549 cell lysate and then linearized, or linearized without methylation, and both methylated and unmethylated DNA was used for both electroporation and natural transformation experiments. Although some antibiotic-resistant colonies grew on most of the plates,
PCR and pigment analyses showed that no transformation or homologous recombination had occurred in these strains.

C.4 Discussion

C.4.1 Chl. phaeobacteroides strain 1549. It is possible that this strain cannot take up exogenous DNA because it produces an extracellular polysaccharide (EPS) that inhibits movement of DNA towards the cellular membrane. Although the production of EPS by Chl. phaeobacteroides strain 1549 has not been studied, the fact that the cell pellets are very loose, especially when cells from very dense cultures are centrifuged, suggests that it does synthesize some sort of extracellular carbohydrate, and this can inhibit transformation efficiency (15).

Transformation efficiency of EPS-producing strains might be improved by finding culture conditions in which EPS production is reduced (15). Alternatively, it might be best to find another brown-colored strain suitable for transformation. The strain Pelodictyon phaeum, which is now growing in culture in the lab, may be suitable. It grows on solid CP medium or liquid CL medium amended with 30 g L⁻¹ NaCl and can use thiosulfate as an electron donor. One important technical detail is that at high salt concentrations, Phytagel, the usual solidifying agent for CP medium, solidifies as soon as the solution has stopped boiling. Since Pld. phaeum requires 3% salt in the medium, Bacto-Agar rather than Phytagel should be used to solidify the high-salt plates. Additionally, growth of Pld. phaeum on media with lower concentrations of salt should be tested.

C.4.2 The platforms. It is possible that bchU was not the best test gene for targeted inactivation. The pigment produced in a bchU mutant strain of Chl. phaeobacteroides strain 1549 would be BChl f, which has never been observed in nature, and this fact cannot be discounted. Perhaps the hypothetical pigment BChl f remains hypothetical because its aggregates in the chlorosomes are malformed. Perhaps the spectral overlap between emission of BChl f aggregates and absorption of CsmA-BChl a in the baseplate
is insufficient for energy transfer, such that cells producing such a pigment would grow very, very slowly and would require very high light intensities. A gene not required for growth under standard laboratory conditions, such as \textit{nifD}, might be a better test case.

Inactivation of \textit{BE1} and the downstream gene, encoding a predicted dehydrogenase, was also attempted. Although perhaps better than the \textit{bchU} platform in the sense that BChl \textit{c}, which should be produced by this mutant, is common in GSB, these constructs were not optimal because they were based on genes that had been amplified from \textit{Chl. phaeobacteroides} strain 266\textsuperscript{T}, not strain 1549. Although the amino acid sequences of most homologous proteins from GSB seem to be almost identical, their nucleotide sequences vary much more (T. Li, personal communication). For this reason, the DNA sequences of these two constructs may not have been sufficiently similar to the wild-type sequences in \textit{Chl. phaeobacteroides} strain 1549 for homologous recombination to occur.

The genes \textit{cruA} and \textit{crtQ} from \textit{Chl. phaeobacteroides} strain 1549 and \textit{cruA} and \textit{cruB} from \textit{Pld. phaeum} have been amplified, cloned, and sequenced. The plasmids pTA-AP, and pTA-QP (with the \textit{Chl. phaeobacteroides} genes) or pTA-AM and pTA-BM (for the \textit{Pld. phaeum} genes) can be used as starting points for making inactivation constructs of any or all of these genes. Partial sequences of these genes have been deposited at GenBank under the accession numbers EF546377 (\textit{Chl. phaeobacteroides} strain 1549, \textit{cruA}), EF546378 (\textit{Chl. phaeobacteroides} strain 1549, \textit{crtQ}), EF546375 (\textit{Pld. phaeum}, \textit{cruA}) and EF546376 (\textit{Pld. phaeum}, \textit{cruB}). It is interesting to note here that biochemical inhibition of lycopene cyclization has a minimal effect on growth rate in both of these species (see Chapter 3), and for this reason either \textit{cruA} or \textit{cruB} should be suitable test genes for targeted inactivation.

\textbf{C.4.3 Antibiotic resistance}. The concentrations sufficient to prevent growth of \textit{Chl. phaeobacteroides} strain 1549 are comparable to or slightly lower than those used for \textit{Chl. tepidum}, with the exception of gentamicin. When the method for chromosomal gene inactivation in \textit{Chl. tepidum} by natural transformation was first published, 100 \(\mu\text{g} \text{ml}^{-1}\) of gentamicin was enough to prevent growth of \textit{Chl. tepidum}, as it is for \textit{Chl. phaeobacteroides} strain 1549. However, that concentration is no longer sufficient, and concentrations between 350 and 450 \(\mu\text{g} \text{ml}^{-1}\) are now routinely used for \textit{Chl. tepidum}. It
seems likely that rapid development of resistance to gentamicin is a trait that may be common to all GSB or at least to the genus *Chlorobaculum* (*Chl. tepidum* and *Chl. phaeobacteroides* strain 1549 are both members of the newly-named genus *Chlorobaculum*; 8), and it is a trait that could be advantageous in future experiments. A gentamicin-resistant strain that can tolerate more than 20 μg gentamicin ml<sup>−1</sup> (the standard concentration used for *E. coli*) can be used in conjugation experiments. In the conjugation experiments described here, the *E. coli* donor cells remained as contaminants until the third passage, as the selection used was to the plasmid only and the *E. coli* cells retaining the plasmid would be resistant. If a recipient strain resistant to gentamicin were used, double selection – with gentamicin as well as the selection for the plasmid – could be used, and the *E. coli* would be strongly selected against.

It may also be useful to note here that the primers that amplify from the *aacC1* cassette also amplify a fragment of a similar but slightly different size from wild-type *Chl. phaeobacteroides* strain 1549. A different primer pair for amplification from this cassette would eliminate false-positive PCR results from gentamicin-resistant putative transformants.

**C.5 Conclusion.** Although to date none of the attempts to inactivate a gene in *Chl. phaeobacteroides* strain 1549 have succeeded, it should not be impossible to develop a genetic system for a brown-colored strain of GSB. Using a different platform and attempting transformation of *Pld. phaeum* instead of *Chl. phaeobacteroides* strain 1549 may improve the odds of success.

**Acknowledgements.** *E. coli* strain S17-1 and plasmid pKJS2 were the gift of Dr. Kenneth Keiler at The Pennsylvania State University. Dr. John Ormerod at the University of Oslo, Norway, provided *Chlorobium phaeobacteroides* strain 1549. Dr. Robert E. Blankenship from Washington University in St. Louis has provided *Pelodictyon phaeum* for future attempts at transformation.
References


Figure Legends.

Figure C.1 Constructs for inactivation of \textit{bchU} in \textit{Chlorobium phaeobacteroides} strain 1549.

Figure C.2 Construct for inactivation of \textit{BE1} in \textit{Chlorobium phaeobacteroides} strain 1549.

Figure C.3 PCR analysis of putative transformants after electroporation. Primers used are indicated above the figure. M, DNA Ladder I from GeneChoice; fragment lengths (in kbp) are indicated on the left side. Templates from samples 1-3 are from \textit{Chl. phaeobacteroides} strain 1549 gentamicin-resistant putative transformants; sample 4 is from wild-type \textit{Chl. phaeobacteroides} strain 1549; samples 5 and 6 are plasmids pCPB-Gm(+) and pCPB-Gm(-), respectively.

Figure C.4. Absorption spectra of methanol extracts of putative transformants after electroporation. Solid black line, wild-type \textit{Chl. phaeobacteroides} strain 1549. Solid gray line, gentamicin-resistant putative transformant. Dashed black line, streptomycin- and spectinomycin-resistant putative transformant.

Figure C.5 PCR analysis of putative transconjugants. Primers used were (A) BCHU-F1/BCHU-R1, (B) om-2s/om-2as, (C) GmR-F/GmR-R, and (D) BCHU-F1/om-2s. Lanes labeled “M” have DNA Ladder I from GeneChoice; fragment lengths (in kbp) are indicated on the left side of (A) and are the same for all samples. Templates W are from wild-type \textit{Chl. phaeobacteroides} strain 1549, PG is plasmid pCPB-GmR(+), PA+ and PA- are plasmids pCPB-A(+) and pCPB-A(-), respectively. Samples 1-11 in A, B, and D are from streptomycin- and spectinomycin resistant colonies; samples 13-23 in A and C are from gentamicin-resistant colonies.
Table C.1 Antibiotic sensitivity of *Chl. phaeobacteroides* strain 1549.

<table>
<thead>
<tr>
<th></th>
<th>Chloramphenicol</th>
<th>Erythromycin</th>
<th>Gentamicin</th>
<th>Streptomycin/ spectinomycin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Little growth$^a$</td>
<td>5</td>
<td>5</td>
<td>50</td>
<td>75/150</td>
</tr>
<tr>
<td>No growth$^b$</td>
<td>10</td>
<td>8</td>
<td>100</td>
<td>150/300</td>
</tr>
</tbody>
</table>

$^a$Highest concentration of antibiotics on which some growth was visible.

$^b$Lowest concentration of antibiotics on which no growth was visible. All concentrations are in µg ml$^{-1}$. 
Figure C.1
APPENDIX D.

UNEXPECTED DIVERSITY AND COMPLEXITY OF THE GUERRERO NEGRO HYPERSALINE MICROBIAL MAT.
Unexpected Diversity and Complexity of the Guerrero Negro Hypersaline Microbial Mat

Ruth E. Ley,1,2† J. Kirk Harris,1,2 Joshua Wilcox,1,2 John R. Spear,1,2‡ Scott R. Miller,3 Brad M. Bebout,4 Julia A. Maresca,3 Donald A. Bryant,5 Mitchell L. Sogin,6 and Norman R. Pace1,2‡

Department of Molecular, Cellular and Developmental Biology, University of Colorado, Boulder, Colorado 803091; Center for Astrobiology, University of Colorado at Boulder, Boulder, Colorado 803092; Division of Biological Sciences, The University of Montana, Missoula, Montana 598123; Exobiology Branch, NASA Ames Research Center, Moffett Field, California 940354; Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, Pennsylvania 168025; and The Marine Biological Laboratory, Woods Hole, Massachusetts 025436

Received 14 December 2005/Accepted 26 February 2006

We applied nucleic acid-based molecular methods, combined with estimates of biomass (ATP), pigments, and microelectrode measurements of chemical gradients, to map microbial diversity vertically on a millimeter scale in a hypersaline microbial mat from Guerrero Negro, Baja California Sur, Mexico. To identify the constituents of the mat, small-subunit rRNA genes were amplified by PCR from community genomic DNA extracted from layers, cloned, and sequenced. Bacteria dominated the mat and displayed unexpected and unprecedented diversity. The majority (1,336) of the 1,586 bacterial 16S rRNA sequences generated were unique, representing 752 species (>97% rRNA sequence identity) in 42 of the main bacterial phyla, including 15 novel candidate phyla. The diversity of the mat samples differentiated according to the chemical milieu defined by concentrations of O2 and H2S. Bacteria of the phylum Chloroflexi formed the majority of the biomass by percentage of bulk rRNA and of clones in rRNA gene libraries. This result contradicts the general belief that cyanobacteria dominate these communities. Although cyanobacteria constituted a large fraction of the biomass in the upper few millimeters (>80% of the total rRNA and photosynthetic pigments), Chloroflexi sequences were conspicuous throughout the mat. Filamentous Chloroflexi bacteria were identified by fluorescence in situ hybridization within the polysaccharide sheaths of the prominent cyanobacterium Microcoleus chthonoplastes, in addition to free living in the mat. The biological complexity of the mat far exceeds that observed in other polysaccharide-rich microbial ecosystems, such as the human and mouse distal guts, and suggests that positive feedbacks exist between chemical complexity and biological diversity.

Microbial mats are benthic aquatic ecosystems fueled by light energy and composed of microbial cells attached to extracellular polymeric material and mineralized scaffolds in visible millimeter scale layers (12). Unlike stromatolites, which have a biotic mechanism for calcification (49), microbial mats become layered because of occasional sedimentation and regrowth. Microbial mats and stromatolites are found in the fossil record dating back 3.4 billion years (60) and are thought to have significantly influenced the composition of the atmosphere with production of O2, H2, and CH4 (21). Ancient and fossil record dating back 3.4 billion years (60) and are thought to have significantly influenced the composition of the atmosphere with production of O2, H2, and CH4 (21). Ancient and fossil record dating back 3.4 billion years (60) and are thought to have significantly influenced the composition of the atmosphere with production of O2, H2, and CH4 (21). Ancient and fossil record dating back 3.4 billion years (60) and are thought to have significantly influenced the composition of the atmosphere with production of O2, H2, and CH4 (21). Ancient and fossil record dating back 3.4 billion years (60) and are thought to have significantly influenced the composition of the atmosphere with production of O2, H2, and CH4 (21). Ancient and fossil record dating back 3.4 billion years (60) and are thought to have significantly influenced the composition of the atmosphere with production of O2, H2, and CH4 (21). Ancient and fossil record dating back 3.4 billion years (60) and are thought to have significantly influenced the composition of the atmosphere with production of O2, H2, and CH4 (21). Ancient and fossil record dating back 3.4 billion years (60) and are thought to have significantly influenced the composition of the atmosphere with production of O2, H2, and CH4 (21). Ancient and fossil record dating back 3.4 billion years (60) and are thought to have significantly influenced the composition of the atmosphere with production of O2, H2, and CH4 (21). Ancient and fossil record dating back 3.4 billion years (60) and are thought to have significantly influenced the composition of the atmosphere with production of O2, H2, and CH4 (21).

The microbial mats within the hypersaline lagoons of the Exportadora de Sal SA saltern in Guerrero Negro, Baja California Sur, Mexico, cover an extensive area of artificial shallow lagoons protected from tidal disturbance by levees. Biogeochemical studies of these mats have shown that oxygen and light, as well as photosynthetic capacity, are rapidly depleted with depth. Degradation of organic matter occurs largely by two processes, fermentation and substrate oxidation through sulfate reduction (12). Both processes take on unusual characteristics in the mats. Fermentation contributes to molecular hydrogen release into the overlying water column, even in the presence of oxygen, such that bubbles of mixed oxygen and hydrogen gases form on the surface of the mat (21). In addition, the highest rates of sulfate reduction occur in the upper, oxygen-rich layers of the mat (5). Known sulfate-reducing bacteria of the delta group of proteobacteria occupy the anoxic zone, however, which suggests that novel groups of sulfate-reducing bacteria reduce sulfate in the mats aerobically (51).

Despite the intriguing biogeochemistry of hypersaline microbial mats and their importance as model systems for studies of the early Earth (21), the composition of the microbiota has not been surveyed comprehensively by culture-independent molecular methods. Classic microbiological studies and limited molecular studies have shown that cyanobacteria dominate the surface layers and revealed five of the other main bacterial phyla (phylogenetic divisions Chloroflexi, Spirochaetes, Proteobacteria, Bacteroidetes, and Firmicutes) and thus indicated a
relatively simple community with little deep evolutionary diversity (10, 39–42, 51, 57). The biological simplicity of the mats was the basis for their recommended use as model systems for metagenomic analyses (4). The dominance of cyanobacteria and the biological simplicity of the community have not been verified by culture-independent methods, however.

The aim of this study was a more comprehensive description of the microbial diversity within the mats and how that diversity is distributed in relation to depth and chemical gradients characteristic of the depth profile. We studied a limited area of one mat intensively. In situ gradients of O₂ and H₂S concentrations and pH, measured on a micrometer scale with microelectrodes, provided a backdrop of vertical chemical gradients onto which we mapped the biological data. ATP concentrations were measured to provide an estimate of living biomass distribution throughout the mat. Pigment concentrations measured by high-pressure liquid chromatography (HPLC) offered a view of the distribution of oxygenic chlorophyll a (Chl a)–containing cyanobacteria in relation to bacteriochlorophyll (BChl)–bearing anoxygenic photosynthetic bacteria. A survey of rRNA genes provided a culture-independent assessment of dominant organisms. For each of 10 layers that divided the mat into a millimeter-centimeter scale depth profile, the composition and diversity of communities were determined by sequence analysis of 16S rRNA genes generated by PCR with universal and bacterium-specific primers from community genomic DNA. We used RNA extraction and quantitative hybridization with group-specific probes as a PCR-independent verification of the abundance of the dominant group identified by sequence analysis and of the cyanobacteria. In addition, we visualized the morphologies and associations of these bacteria by fluorescence in situ hybridization (FISH) with tyramide signal amplification to overcome the intrinsic fluorescence of the mat. These studies collectively revealed unexpected diversity, complexity, and structure within the mat.

MATERIALS AND METHODS

Sample collection. We studied the microbial mat underlying pond 4 (near 5) of the Exportadora de Sal SA, a solar saltworks located at Guerrero Negro, Baja California Sur, Mexico (see reference 38 for site details). The mat is covered by ~1 m of brine with a salinity of ~80‰. Samples were collected in June and October of 2001 at 4 a.m. (night) and 1 p.m. (day). We collected replicate cores HH11011 of the Exportadora de Sal SA, a solar saltworks located at Guerrero Negro, Baja California Sur, Mexico, distributed in relation to depth and chemical gradients measured by high-pressure liquid chromatography (HPLC) offered a view of the distribution of oxygenic chlorophyll a (Chl a)–containing cyanobacteria in relation to bacteriochlorophyll (BChl)–bearing anoxygenic photosynthetic bacteria. A survey of rRNA genes provided a culture-independent assessment of dominant organisms. For each of 10 layers that divided the mat into a millimeter-centimeter scale depth profile, the composition and diversity of communities were determined by sequence analysis of 16S rRNA genes generated by PCR with universal and bacterium-specific primers from community genomic DNA. We used RNA extraction and quantitative hybridization with group-specific probes as a PCR-independent verification of the abundance of the dominant group identified by sequence analysis and of the cyanobacteria. In addition, we visualized the morphologies and associations of these bacteria by fluorescence in situ hybridization (FISH) with tyramide signal amplification to overcome the intrinsic fluorescence of the mat. These studies collectively revealed unexpected diversity, complexity, and structure within the mat.

sequence analysis of 16S rRNA genes generated by PCR with universal and bacterium-specific primers from community genomic DNA. We used RNA extraction and quantitative hybridization with group-specific probes as a PCR-independent verification of the abundance of the dominant group identified by sequence analysis and of the cyanobacteria. In addition, we visualized the morphologies and associations of these bacteria by fluorescence in situ hybridization (FISH) with tyramide signal amplification to overcome the intrinsic fluorescence of the mat. These studies collectively revealed unexpected diversity, complexity, and structure within the mat.

sequence analysis of 16S rRNA genes generated by PCR with universal and bacterium-specific primers from community genomic DNA. We used RNA extraction and quantitative hybridization with group-specific probes as a PCR-independent verification of the abundance of the dominant group identified by sequence analysis and of the cyanobacteria. In addition, we visualized the morphologies and associations of these bacteria by fluorescence in situ hybridization (FISH) with tyramide signal amplification to overcome the intrinsic fluorescence of the mat. These studies collectively revealed unexpected diversity, complexity, and structure within the mat.

sequence analysis of 16S rRNA genes generated by PCR with universal and bacterium-specific primers from community genomic DNA. We used RNA extraction and quantitative hybridization with group-specific probes as a PCR-independent verification of the abundance of the dominant group identified by sequence analysis and of the cyanobacteria. In addition, we visualized the morphologies and associations of these bacteria by fluorescence in situ hybridization (FISH) with tyramide signal amplification to overcome the intrinsic fluorescence of the mat. These studies collectively revealed unexpected diversity, complexity, and structure within the mat.
into operational taxonomic units (OTUs) by pairwise identity (% ID) with a furthest-neighbour algorithm and a precision of 0.01 implemented in DOTUR (52). We used DOTUR to determine OTU frequencies in mouse cecal and human colonic 16S rRNA gene sequence data sets with Arb alignments provided by the authors (15, 32). Simpson's diversity index and collector's curves were calculated with EstimateS (6).

Assignment of the majority of sequences to their respective phyla was based on their position after parsimony insertion into the Arb dendrogram (omitting hypervariable portions of the rRNA gene with the lanemaskPH provided with the database; see http://Pacelab.colorado.edu/Publications/publications.html). Sequences that did not fall within described phyla were further characterized. Phylogenetic trees including the novel sequences and reference taxa (20) were constructed by evolutionary distance (test version 4.0.2 of PAUP*, a neighbor-joining algorithm with either Kimura two-parameter correction or maximum-likelihood [ML] correction with an empirically determined gamma distribution model of site-to-site rate variation and empirically determined base frequencies), parsimony (test version 4.0.2 of PAUP*), heuristic search, and ML (fastDNAm) analyses. Bootstrap resampling was used to test the robustness of inferred topologies. Novel candidate phyla, designated GN01 to GN15, were defined by the generally accepted criteria that (i) there must be three or more sequences from independent PCR products, (ii) sequences must be a minimum of 1,000 bp, and (iii) there must be high levels of support in phylogenetic analyses (20, 24, 46). However, four of the novel candidate phyla (GN6, GN12, GN13, and GN14) met only two of these criteria (the sequences were <1,000 bp) and await confirmation with longer sequence reads.

To cluster the communities from each layer, we used the UniFranc computational tool (35). The Arb alignment (excluding hypervariable regions) containing all the novel sequences was used to construct an ML tree with RAxML (54). The tree was annotated according to the layer from which each sequence was derived, and the fraction of tree branch length unique to any one layer in pairwise comparisons (the UniFranc metric) was calculated. Microbial communities from individual layers were clustered by application of the unweighted-pair group method using average linkages (UPGMA) to the UniFranc metric matrix.

**FISH of Chloroflexi.** Mat samples were disrupted gently with a pestle, dehydrated in ethanol, and adhered to silane-coated glass slides with Cell-Tak (BD Biosciences, Inc.). Cells were permeabilized with lysozyme (10 mg/ml) 30 min, 37°C), followed by achromopeptidase (60 U/ml, 1 h, 37°C) and mutalysin (10×, 30 min, 37°C) (Sigma Aldrich, Inc.). Hybridizations were performed with a buffer containing 30% formamide, 0.9 M NaCl, 20 mM Tris-HCl (pH 8.0), 0.01% sodium dodecyl sulfate, and 50 ng of biotin-labeled probe for 2 h at 46°C. Signal intensity was boosted with TSA kit no. 22 with horseradish peroxidase-streptavidin and Alexa Fluor 488 tyramide by following the manufacturer’s instructions (Molecular Probes, Carlsbad, CA). The probes used in this study were Chloroflexi-specific Chloroflexi9415′-AAA CCA CAC GCT CCG CT-3′ (18) and bacterium-specific EUB3385′-GCT TCC TCG AGT AGT-3′ (1). Samples were counterstained with 10 μg of 4′,6-diamidino-2-phenylindole (DAPI)/ml and mounted with antifade (CitiFluor Ltd., Leicester, England).

Images were generated by laser confocal microscopy (Leica Microsystems, Barrington, IL) with a 488-nm excitation laser and a 350-nm excitation laser (DAPI)/ml and mounted with antifade (CitiFluor Ltd., Leicester, England). Probes and protocols were tested against reference strains (20) with a 488-nm excitation laser and a 350-nm excitation laser (DAPI). Probes and protocols were tested against reference strains (20) with a 488-nm excitation laser and a 350-nm excitation laser (DAPI). Probes and protocols were tested against reference strains (20) with a 488-nm excitation laser and a 350-nm excitation laser (DAPI). Probes and protocols were tested against reference strains (20) with a 488-nm excitation laser and a 350-nm excitation laser (DAPI). Probes and protocols were tested against reference strains (20) with a 488-nm excitation laser and a 350-nm excitation laser (DAPI). Probes and protocols were tested against reference strains (20) with a 488-nm excitation laser and a 350-nm excitation laser (DAPI). Probes and protocols were tested against reference strains (20) with a 488-nm excitation laser and a 350-nm excitation laser (DAPI). Probes and protocols were tested against reference strains (20) with a 488-nm excitation laser and a 350-nm excitation laser (DAPI). Probes and protocols were tested against reference strains (20) with a 488-nm excitation laser and a 350-nm excitation laser (DAPI).

**RESULTS**

**Depth gradients of H2S, O2, and pH.** The in situ chemical environment of the mat was profiled by microelectrode measurements of O2 and H2S concentrations and of pH from the surface of the mat to a depth of 6 mm in 200-μm increments. The high spatial resolution of these measurements provides a view of the variability of the local chemistry. Based on our microelectrode measurement profiles, the mat can be divided into three distinct habitats that serve as a backdrop for the spatial organization of the microbial community (Fig. 1A). These zones are the oxic zone, ranging from the top of the mat to a depth of 2 mm and characterized by diurnally fluctuating concentrations of O2; the low-H2S zone, ranging in depth from 2 mm to 6 mm, where H2S levels are drawn down diurnally; and the H2S-rich zone, the largest zone, ranging from 6 mm to the bottom (~60 mm), where concentrations of H2S are permanently high (Fig. 1A). The coefficient of variation for the O2 measurements was greatest at an average depth of 2.5 mm, where the average O2 concentration was low (28 μM; Fig. 1A). This is an indication that the deepest penetration of oxygen into the mat is also the depth at which oxygen concentrations are the most variable (measured range, 0 to 522 μM) during the day. At night, the mat is completely anoxic.

**Biomass and pigment distributions by depth.** To determine the depth distribution of biomass in the mat, we measured the concentration of ATP per gram, from the top of the mat to the bottom, in June and October. ATP is the energy currency of all cells, and it therefore serves as a proxy for biomass (28). ATP concentrations were highest in the oxic zone (1,343 ± 400 ng of ATP g-1; Fig. 1B) and tapered off rapidly with depth to 322 ± 50 ng of ATP g-1 in the lower, H2S-rich zone. Integrating by depth, the substantially greater volume of the lower H2S-rich zone results in fivefold greater overall biomass per unit of surface area: 27.8 μmol of ATP cm-2 in the lower zone versus 5.0 μmol of ATP cm-2 for the combined oxic and low-H2S zones. The ATP concentrations and depth profiles of the October and June sampling dates were similar.

We determined the concentrations and distributions of photosynthetic pigments by HPLC. Pigment concentrations were...
highest in the oxic zone, as expected, yet we detected pigments at all depths (Fig. 1C). The cyanobacterial pigment Chl a was an order of magnitude more abundant than BChls (BChls a, d, and c) in the oxic zone.

Estimates of richness and coverage. Two rRNA gene clone libraries were constructed from each of the 10 layers, one with bacterium-specific primers (8F and 1391R) and one with universal primers (515F and 1391R). Restriction fragment length polymorphism screening revealed unexpectedly high levels of diversity for each 96-clone library; therefore, all clones were sequenced bidirectionally. When the universal libraries from all layers were combined, the bacteria/archaea/eucarya ratio was 57:7:1. In this report, we focus on the bacterial sequences. Archaeal and eucaryal sequence data will be described elsewhere (J. Spear, R. Ley, and N. R. Pace, unpublished). Bacterial sequences encountered in the libraries constructed with archaeal primers were included in the final data set discussed here. The two forward primers (515F and 8F) used to generate the majority of the sequences yielded an equivalent proportion of bacterial phyla, an indication that our coverage was not significantly biased by the primer pairs that we used. For subsequent analyses, all sequences were combined on the basis of the layer from which they originated.

To assess the coverage and the richness of clone libraries combined by layer, we employed collector’s curves for coverage, the nonparametric estimators Chao1 and Ace1 for richness, and the computed Simpson index to estimate the evenness of community composition (see reference 52). The total bacterial 16S rRNA gene sequences (n = 1,586) consisted of 1,336 unique sequences and 752 phylotypes defined by a minimum threshold of 99% ID. In all subsequent analyses, we report masked sequence pairwise identities (% IDs). The hypervariable regions of sequences were masked for alignment purposes because these regions cannot be aligned with certainty, particularly across the large phylogenetic distances encountered in this data set. Generally, 99% ID is equivalent to ~97% ID when the entire sequence length is used (e.g., see the data set of 11,831 bacterial 16S rRNA genes in reference 15), often taken to indicate species level variation. Figure 2A shows the frequency of observed taxa with % ID thresholds ranging from 90% to 100%. The Chao1 and the Ace1 richness estimates for phylotypes ranging from 90% ID to 100% ID yielded equivalent richness curves (Fig. 2A). Both richness estimators indicated >10,000 unique sequences based on the distribution of observed sequences. Collector’s curves for taxa with >90% ID indicate that our coverage of the diversity was not comprehensive, since the curves did not begin to become asymptotic (Fig. 2B). Together, the richness estimators and the collector’s curves indicate a high degree of diversity, most of which likely remains undescribed. Bacterial diversity was uneven in the upper layers sampled, which had a subset of dominant bacteria, and became more even with depth (Simpson’s index versus depth, R^2 = 0.63, P < 0.005).

Bacterial diversity and distribution. The phylum Chloroflexi dominated clone libraries numerically and included the highest proportion of sequences, on average (Fig. 3A), from each chemically defined zone (Fig. 3B, C, and D). Proteobacteria and Bacteroidetes were the second most represented phyla in each zone. Cyanobacterial sequences comprised, at most, 10% of
the total sequences in a chemical zone (Fig. 3B) and were obtained from the oxic zone only. Figure 4 shows the phylogenetic relationships of the Chloroflexi and cyanobacterial sequences obtained from the mat in relation to cultured representatives and clones obtained by culture-independent methods from other environments. Twenty-four percent of the Chloroflexi sequences were obtained from the surface layers and were close relatives of known photosynthetic organisms, such as Chloroflexus and Chlorothrix spp., previously described for these mats (30, 41) (Fig. 4A). Another group of Chloroflexi sequences (4.5% of Chloroflexi sequences) with no cultured close relative was obtained only from the surface layers (Fig. 4A). This result indicates that the Chloroflexi may include additional unrecognized photosynthetic members. The other Chloroflexi groups recovered from the mat had average depth distributions extending below the oxic zone. These groups include oxygen-tolerant members, however, since members of all groups were obtained from the oxic zone. The majority of sequences have no closely related cultured representatives from which properties can be inferred. Sixty-one percent of the cyanobacterial sequences obtained from the mat were members of a novel group of cyanobacteria whose closest relative was obtained from a hot spring mat (Fig. 4B). Sixteen percent formed a group affiliated with Leptolyngya spp.

In addition to Chloroflexi and Cyanobacteria, the 16S rRNA gene sequence analysis revealed 28 other previously described phyla (Fig. 3). Half of these were candidate phyla, so termed because they are known only by 16S rRNA gene sequences and do not contain representatives that have been cultured in the laboratory and from which physiological and metabolic properties can be inferred. This study expanded the known habitat space and diversity of several candidate phyla. For example, candidate phylum KSB1, previously represented by 110 sequences encountered in surveys of brackish water sediments (59), cave sediments (22), and a bioreactor treating 4-methylbenzoate (62), was expanded significantly with the addition of 35 sequences from the mat.

The majority of sequences could be assigned unambiguously to known phyla; however, 119 sequences remained (7.5% of the total) that were not affiliated with any known phylum. Phylogenetic analysis of these unaffiliated sequences revealed 15 novel candidate phyla termed GN01 to GN15. Several of the candidate phyla include sequences that were previously deposited in GenBank and for which there is no described

FIG. 3. Bacterial diversity in the mat. Proportions of bacterial phyla in the total data set (A), in the oxic zone (0 to 2 mm) (B), in the low-H2S zone (2 to 6 mm) (C), and in the H2S-rich zone (6 to 60 mm) (D) are shown. Others: cyanobacteria, KSB1, OP10, OP3, OP5, GN1, Firmicutes, OP11, GN04, GN05, GN09, GN10, WS1, WS2, GN2, Deinococcus-Thermus, GN07, Haloanaerobiales, GN06, GN11, BRC1, OP8, OS-K, GN12, GN13, GN14, actinobacteria, GN15, WS3, GN8, OP9, TM6, and VadinBE97. Abbreviations: Chloro., Chloroflexi; Cyano., Cyanobacteria; Verr., Verrucomicrobia; Planct., Planctomycetes; Spiro., Spirochaetales; Firm., Firmicutes, Bact., Bacteroidetes; Proteo., Proteobacteria. (E) Bacterial community clustering by layer studied (UPGMA tree of UniFrac metric based on 1,585 16S rRNA gene sequences). Shaded areas refer to the different chemical milieus identified by the microelectrode measurements in Fig. 1A.
A. Chloroflexi

Most of the novel GN candidate phyla include sequences derived from several different layers and therefore several separate PCR products. Novel GN candidate phyla were detected in all layers of the mat, and more than half were detected in the oxic zone (Table 1).

Figure 5 shows the depth distribution of members of the 14 most commonly observed bacterial phyla. The abundance of each is shown as the percentage of sequences at each depth, calculated as a fraction of the total number of sequences obtained for any given phylum. The Chloroflexi bacteria are distributed fairly evenly with depth, except for a relative reduction in abundance in the transitional low-H$_2$S zone. The phyla that were most abundant in the oxic zone included Cyanobacteria, Proteobacteria, Bacteroidetes, Spirochaetes, Verrucomicrobia, and candidate phylum GN01. Several known candidate phyla (KSB1, OP10, OP5, and OD1) had their highest abundances in the low-H$_2$S zone, just below the oxic layer. Firmicutes exhibited a marked bimodal distribution, abundant in the oxic zone and in the lower portion of the mat.

B. Cyanobacteria

Community similarities. In order to compare the communities within the chemical zones characterized by the microelectrode measurements of O$_2$ and H$_2$S, we used the recently developed UniFrac metric analysis (35). UniFrac measures the phylogenetic distance between pairs of communities represented by sequences in a phylogenetic tree as the fraction of branch length of the tree that leads to descendants from either one community or the other but not both (35). A phylogenetic affiliation (Table 1; a dendrogram is available in the Arb database at http://Pacelab.colorado.edu/Publications/publications.html) for the Arb dendrogram showing these phyla in the context of previously described phyla.

### TABLE 1. Mat layers from which sequences forming candidate phyla GN01 to GN15 were obtained

<table>
<thead>
<tr>
<th>Phylum</th>
<th>Depth(s) in mm (no. of sequences)</th>
<th>Accession no. of clones from other studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>GN14</td>
<td>0–1 (2)</td>
<td></td>
</tr>
<tr>
<td>GN07</td>
<td>0–1 (2), 3–4 (1), 4–5 (1)</td>
<td></td>
</tr>
<tr>
<td>GN03</td>
<td>0–1 (2), 2–3 (1), 4–5 (7), 5–6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(9), 6–13 (2), 13–26 (5)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td></td>
</tr>
<tr>
<td>GN01</td>
<td>1–2 (8), 2–3 (1), 3–4 (6), 4–5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1), 5–6 (6)</td>
<td></td>
</tr>
<tr>
<td>GN05</td>
<td>1–2 (1), 2–3 (1), 3–4 (3), 4–5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4), 5–6 (2), 26–39 (1), 39–60 (1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1), 26–39 (6)</td>
<td></td>
</tr>
<tr>
<td>GN08</td>
<td>3–4 (1)</td>
<td></td>
</tr>
<tr>
<td>GN11</td>
<td>3–4 (1), 4–5 (1), 5–6 (1)</td>
<td></td>
</tr>
<tr>
<td>GN02</td>
<td>3–4 (2), 4–5 (2), 13–26 (1)</td>
<td></td>
</tr>
<tr>
<td>GN12</td>
<td>4–5 (1), 5–6 (1)</td>
<td></td>
</tr>
<tr>
<td>GN13</td>
<td>5–6 (1), 6–13 (1)</td>
<td></td>
</tr>
<tr>
<td>GN06</td>
<td>5–6 (2), 26–39 (1)</td>
<td></td>
</tr>
<tr>
<td>GN04</td>
<td>5–6 (3), 6–13 (1), 13–26 (3)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>26–39 (4), 39–60 (2)</td>
<td></td>
</tr>
<tr>
<td>GN15</td>
<td>26–39 (2)</td>
<td></td>
</tr>
</tbody>
</table>

*Values indicate the depths in the mat from which candidate phylum clones were obtained; the number of sequences from each depth is in parentheses. The GenBank accession numbers are those of clones from other studies without prior phylum affiliations that are included in GN candidate phyla. See http://Pacelab.colorado.edu/Publications/publications.html for the Arb dendrogram showing these phyla in the context of previously described phyla.*
Microbial mats occur worldwide in shallow aquatic environments where high salinity or temperature precludes the establishment of algae or aquatic plants that may overgrow the mats or of grazers that may otherwise consume them (17). Prior to the evolution of such organisms, microbial mat-like structures were widespread around shallow seas and lakes and are thought to have contributed significantly to the evolution of Earth’s atmosphere (21). The biogeochemical cycling of carbon, oxygen, and hydrogen has been studied extensively in the hypersaline mats of Guerrero Negro, Baja California Sur, Mexico (5, 12, 13, 61), yet the identities of the organisms that drive these cycles are known primarily from culture and microscopy studies (9) and only limited molecular analyses have been performed (39–41, 51). This study is the most extensive rRNA-based survey of a microbial mat so far conducted. Furthermore, we integrated a variety of molecular and chemical analytical approaches to characterize the depth profile of the mat. In contrast to the long-held view that these mats are dominated by cyanobacteria and are biologically simple, our results indicate an overall unexpected dominance of *Chloroflexi* bacteria and a remarkably high level of diversity. We found that cyanobacteria dominate the biomass in the upper 2 mm;
however, the dominance of *Chloroflexi* bacteria below 2 mm, where the majority of the biomass resides, makes them the most abundant type of bacterium overall. In the upper 2 mm, the *Chloroflexi* bacteria can be seen intertwined within the exopolysaccharide sheaths of the dominant cyanobacterium. Together, our results indicate that the photosynthetic activity of the cyanobacteria sustains a highly diverse and structured community.

**Chemically defined niches and biomass distribution.** Our microelectrode measurements of H$_2$, O$_2$, and pH provided a detailed view of the chemical environment in the upper 6 mm of the mat. These chemical profiles defined three distinct chemical niche spaces onto which to map the biological diversity of the mat. These three zones, the oxic zone, the low-H$_2$S zone, and the H$_2$S-rich zone, have been described previously with coarser spatial resolution (5).

In previous studies, the density of the cells was observed qualitatively to be highest in the upper few millimeters and to taper off with depth (10). Bulk RNA concentrations have also been shown to be higher by an order of magnitude in the upper 1 mm than in the underlying layers (51). We measured ATP concentrations in the mat as a proxy for biomass. ATP concentrations approached those of pure cell paste in the uppermost millimeters of the mat. The upper few millimeters are considered to be the highly active zone of the mat, where carbon flux driven by photosynthesis is highest. However, even though ATP concentrations were lowest below the oxic zone, the total amount of ATP present per unit of area in the lower portion of the mat is highest below the oxic zone. Therefore, more overall biomass resides in the dark sulfidic part of the mat than in the cell-dense upper few millimeters, where most photosynthesis occurs.

Pigment depth profiles showed that the cyanobacterial pigment Chl $a$ was by far the most abundant photosynthetic pigment. It is therefore likely that cyanobacteria fix more inorganic C than the nonoxyngeic phototrophic bacteria (e.g., *Chloroflexi* bacteria and proteobacteria) and thus the cyanobacteria presumably provide the main sustenance of the mat, as previously thought. We measured peak concentrations of the BChls directly under the peak concentration of Chl $a$. This segregation with depth is consistent with the specialization of BChls for longer, more deeply penetrating wavelengths and the previously reported vertical stratification of different phototrophic organisms (10, 56). The photosynthetic *Chlo-

![FIG. 6. Chloroflexi bacteria and the cyanobacterium *M. chthonoplastes* in the mat visualized by laser confocal microscopy. (A) *Chloroflexi* bacteria (red, FISH *Chloroflexi* probe) entwined with *M. chthonoplastes* (green, DAPI) at a 1-mm depth. The arrow indicates the edge of the polysaccharide sheath. (B) *Chloroflexi* bacteria (green, *Chloroflexi* probe) and *M. chthonoplastes* (green, autofluorescence of Chl $a$). (C) *Chloroflexi* bacteria (thin filaments) and *M. chthonoplastes* (thick filaments), DAPI stained. (D) *Chloroflexi* filaments (red, *Chloroflexi* probe) and polysaccharide material (dull green) at a 50-mm depth. Non-*Chloroflexi* bacteria are visible as bright green spots (arrow 1). Arrow 2 indicates a buried *M. chthonoplastes* filament. Scale bars, 10 μm.
rofexi bacteria (e.g., Chlorothrix halophila and relatives), which are similar in properties to green sulfur bacteria by having chlorosomes, may also contribute in a substantial way to the primary productivity of the mat. Taken together, the biomass and pigment depth distributions indicate that photosynthetic activity within theoxic zone fuels the underlying, larger biomass. Such an inverted pyramid model of trophic levels, where consumer biomass apparently outweighs producer biomass, has recently been described for the stromatolites (calcified mats) of Shark's Bay, Australia (44) and may be common in microbial ecosystems where primary producers exude polysaccharides.

**Diversity mapped onto chemical gradients.** We generated 1,586 bacterial 16S rRNA gene sequences from the entire depth profile of the mat, although this number appears to undersample woefully the diversity present in the mat. Together, the diversity estimates based on our sequence coverage and the collector's curves all indicate a highly diverse community, most of which still awaits description. Despite the low sampling coverage, the high proportion of known and new bacterial phyla in this hypersaline mat makes it the most biologically diverse environment yet characterized. This study has encountered the highest number of confirmed and potential candidate phyla so far observed in a single environment (25). This is remarkable, particularly at a time when the discovery rate of new phyla was thought to be tapering off (46), and suggests that diversity and complexity are a feature of the microbial mat ecosystem.

Previous microscopy-based studies have indicated a vertical stratification of the microbial community with depth, down to about 6 mm, the maximum depth studied (10). Cyanobacteria, particularly the filamentous, exopolysaccharide sheath-forming species *M. chthonoplastes*, were observed by microscopy in the upper 2 mm. Other, anoxygenic photosynthetic bacteria such as *Chloroflexus*-like *Chloroflexi* bacteria were observed below the cyanobacteria along with other undefined microbiotas with various morphologies (10). We found cyanobacterial rRNA genes in the uppermost layers only, and the majority of the RNA in the upper few millimeters was cyanobacterial. Furthermore, cyanobacterial Chl a was far more abundant in the upper few millimeters than BCHls c and d, ascribed to the *Chloroflexi* bacteria. The discrepancy between *Chloroflexi* bacterial and cyanobacterial representation in 16S rRNA gene libraries versus bulk rRNA and pigment profiles in the upper millimeters could be due to the comparatively large size of cyanobacteria. For instance, *M. chthonoplastes* cells are two to three times wider than *Chloroflexi* bacterial cells, which may result in a higher ratio of expressed RNA to rRNA gene copy number. Nonetheless, overall, the *Chloroflexi* bacteria dominated the mat biomass, with the exception the top 2 mm, where cyanobacteria were a larger proportion of the biomass. Popular reference to the mats as “cyanobacterial” reflects the focus of most previous work on the top few millimeters.

The majority of the novel candidate phyla that we describe from the mat were detected at a variety of depths. Candidate phyla are known from their constituent 16S rRNA gene sequences alone, and therefore the physiological attributes of the bacteria within them can be gleaned only from the context in which they were discovered. The depth distributions of the novel GN candidate phyla provide testable hypotheses about the physiologies of the organisms. More importantly, the spatial distribution of the candidate phyla can help direct efforts to bring representatives into culture for physiological studies. The middle, low-H$_2$S zone harbored 9 of the 15 GN candidate phyla, indicating that such organisms are anaerobic and H$_2$S tolerant yet may rely on diurnal variation in H$_2$S levels and the H$_2$S-O$_2$ interface. In contrast, candidate phyla GN4 and GN12 were detected only well below the oxic layer and are likely to be composed of strict anaerobes. GN14 was observed only in the top layer (day and night libraries), which suggests that it could be a novel photosynthetic group.

The three distinct zones delineated by concentration gradients of O$_2$ and H$_2$S harbored distinct bacterial communities. The UniFrac analysis of the phylogenetic tree representing the entire data set resulted in clusters of the 10 separate samples according to the depth distributions of the three chemical habitats delineated by our in situ microelectrode measurements. This observation implies that related bacteria occupy similar chemical niches. Therefore, despite the potential for horizontal gene transfer to confer many physiological traits on distantly related bacteria, phylogenetic groups appear to share physiological properties that are manifested as chemical niche preferences at particular depths in the mat. For many of the novel phylotypes observed, the depth at which they were found is consistent with what is known about the physiologies of their close relatives. For instance, close relatives of known photosynthetic organisms, such as *Chloroflexi* relatives of the photosynthetic *Chloroflexus* and *Chlorothrix* spp. previously described for these mats (30, 41), were abundant in the upper zone, where light penetrates. Similarly, close relatives of known sulfate-reducing members of the delta group of proteobacteria were most abundant in the oxic zone, which also is where rates of sulfate reduction are highest (39). Members of the phylum *Bacteroidetes*, known to degrade polysaccharides anaerobically (63), were abundant throughout the dark anoxic zone.

**Chloroflexi bacteria and cyanobacteria: symbiosis or antagonism?** A previously described feature of hypersaline microbial mats is the close physical interaction between the cyanobacterium *M. chthonoplastes* and a thinner filamentous bacterium also inside the polysaccharide sheath. The thinner partner was thought to be a member of either *Proteobacteria* or *Chloroflexi* based on physiological properties and transmission electron microscopy (9). We used FISH with *Chloroflexi* bacterium-specific probes and laser confocal microscopy to confirm that the filamentous bacteria inside the *M. chthonoplastes* sheaths were * Chloroflexi* bacteria (Fig. 6A to C). The association is most often observed at 0.3 to 1.2 mm (5, 9), corresponding to the zone below the region of maximal oxygenic photosynthesis. Indeed, lower in the mat, *Chloroflexi* bacteria are free living (Fig. 6D). The depth location of the association and physiological experiments (9) have suggested a cometabolism of sulfur: the *Chloroflexi* bacteria may draw down levels of H$_2$S stressful for the cyanobacterium (38), which excretes organic carbon used by the *Chloroflexi* bacteria. However, we observed that *M. chthonoplastes* filaments were often disrupted when *Chloroflexi* bacteria were present (Fig. 6B), suggesting the alternative view that *Chloroflexi* bacteria may parasitize the cyanobacterium under H$_2$S stress. The tight physical association of these bacteria from deeply divergent lineages is an example of...
Microbial diversity. Comparisons of diversity levels between complexity in the mat. The physical and chemical microbial interactions that build accretion of community structures important in mats. Thus, the use of particular wavelengths of light (6, 7, 27, 50, 56). Addition-wavelengths creates a stratification of phototrophs adapted to the gut is phototrophy: the differential depth penetration of light moieties (61). Another niche available in mats but not in the distal exploitation by the microbiota of many intermediate chemical to sulfate. The prominence of the sulfur cycle in mats allows the proportion of electrons produced by diagenesis is likely shunted where sulfide production levels are high, however, a larger bacteria (53). Fermentation dominates organic matter diagen-form scaffolds that provide attachment sites and nutrients for microbiota. The mammalian distal gut, like microbial mats, is due to the far greater diversity of chemical niches in the mat, structure between these two microbial systems is most likely phylogenetic levels (Fig. 7). The difference in phylogenetic microbiota of the microbial mat is far more diverse at all (more closely related, e.g., at the species-strain level) (2), the ages that diversified at phylogenetically “shallow” levels diversity in the distal gut is characterized by a few deep lin-eages that diversified at phylogenetically “shallow” levels (more closely related, e.g., at the species-strain level) (2), the microbiota of the microbial mat is far more diverse at all phylogenetic levels (Fig. 7). The difference in phylogenetic structure between these two microbial systems is most likely due to the far greater diversity of chemical niches in the mat, which allow more opportunities for specialization within the microbiota. The mammalian distal gut, like microbial mats, is an energy-rich microbial ecosystem where polysaccharides form scaffolds that provide attachment sites and nutrients for bacteria (53). Fermentation dominates organic matter diagon-esis in both systems, although sulfate reduction and methano-genesis also occur in both systems (12, 43). In microbial mats where sulfide production levels are high, however, a larger proportion of electrons produced by diagenesis is likely shunted to sulfate. The prominence of the sulfur cycle in mats allows the exploitation by the microbiota of many intermediate chemical moieties (61). Another niche available in mats but not in the distal gut is phototrophy: the differential depth penetration of light wavelengths creates a stratification of phototrophs adapted to the use of particular wavelengths of light (6, 7, 27, 50, 56). Addi-tionally, the frequent washout inherent to the gut undoubtedly is a powerful selection force against slow-growing microbes and the accretion of community structures important in mats. Thus, the relative complexity of the microbiota in microbial mats probably correlates with broad niche space. Furthermore, biological diver-sity itself can drive diversity through niche creation (16); there-fore, chemical and biological diversity is expected to form positive feedback loops.

Microbial mats have been called simple systems based on microscopy and culture studies. However, our molecular anal-ysis revealed that the hypersaline microbial mats of Guerrero Negro harbor the most complex bacterial assemblage docu-mented to date in any environment, with 42 phyla, including 15 novel candidate phyla. Microbial mats are hot spots of bacte-rial diversity and constitute a rich reservoir of gene diversity for future studies of bacterial evolution and genomic diversity.

ACKNOWLEDGMENTS

We thank D. DesMarais, A. Dalby, L. Angenent, S. Hanada, A. St. Amand, and J. Gordon for valuable assistance and the Exportadora de Sal of Guerrero Negro, Baja California Sur, Mexico, for access to sites. R. Ley was supported in part by an NRC-NASA Astrobiology In-stitutes postdoctoral associationship, and J. Spear was supported by an Agouron Institute postdoctoral fellowship. This work was supported by the NASA Cooperative Agreement with the University of Colorado Center for Astrobiology to N. R. Pace.

REFERENCES


15. Eckburg, P. B., E. M. Bik, C. N. Bernstein, E. Purdom, L. Dethlefsen, M.

FIG. 7. Phylogenetic structure of microbial mat (n = 1,585; this study), human colonic (n = 11,831) (15), and mouse cecal (n = 5,088) (32) 16S rRNA gene sequence data sets. Sequences were clustered into phylotypes based on percent sequence identity (OTUs with similarity thresholds ranging from 65% ID to 100% ID). The ratio of phylotypes at each threshold to the total sequence in each data set is plotted.
DIVERSITY IN A HYPER SALINE MICROBIAL MAT


17. Garrett, P. 1970. Phototrophic consortia: noncompetitive ecological re-


Appendix E

Identification of a homoserine lactonase in *Sulfolobus solfataricus*
Abstract

The bacterial system of homoserine-lactone (HSL)-based intercellular signaling allows the bacteria that use it to determine the population size and coordinate the activities of genes in an entire population. In recent years, a variety of mechanisms for disrupting this signaling system have been identified in both eukaryotes and bacteria. The AiiA proteins first described in Bacillus sp. are zinc-dependent lactonases that hydrolyze the lactone. Homologs of AiiA have been identified in the genomes of four archaeal species, including the acidophilic, thermophilic species Sulfolobus solfataricus strain P2. Here, it is shown that S. solfataricus can break down HSLs and that E. coli cells expressing AiiA from S. solfataricus gain the ability to break down HSLs. This is the first reported homoserine lactonase activity in an archaeon, as well as the first thermostable AiiA.
E.1 Introduction

Quorum sensing using homoserine lactone (HSL) signals has been well-described in a variety of proteobacterial species (for reviews, see 24, 37, 44). At a sufficiently high concentration of HSLs, the HSLs bind to intracellular receptors in the whole population, which activate or repress genes responsible for activities such as extracellular polysaccharide production (4, 14), secretion apparatus (13, 43), biofilm formation (12, 19, 34, 45), or luminescence, the system in which quorum sensing, or autoinduction, was first described (39).

More recently, a variety of systems in other bacterial or eukaryotic species have been described which interfere with HSL-based signaling have been described in both bacteria and eukaryotes. For example, the marine alga Delisea pulchra, synthesizes halogenated furanones that bind to HSL receptors and promote degradation of the protein-signal complex (32, 33). Thus, the bacteria that would otherwise colonize the leaves of the alga are unable to communicate with each other. Another alga, Ulva intestinalis (formerly Enteromorpha), has a planktonic larval stage and settles permanently only on surfaces which have been conditioned by HSL-producing bacteria (29, 42, 47). Some species of Variovorax have been isolated which use an acylase to break down HSL signals, which they can then use as a source of cellular carbon (26, 27). Finally, several species use a zinc-dependent lactonase to hydrolyze the lactone ring of HSLs (7, 17, 30), thus rendering them useless as signaling molecules (Figure E.1).

These lactonases, named AiiA, were first characterized in several Bacillus species (15, 16, 17, 30), and it was shown that plant cells expressing the protein were resistant to infection by pathogenic Erwinia strains (16, 17, 18, 38). Later, some Agrobacterium species were shown to have two AiiA homologs, AttM and AiiB, which are used to regulate HSL signaling in vivo (7, 51, 52). Other HSL-lactonases have been characterized from Klebsiella pneumoniae and Arthrobacter sp. (41). A database search for AiiA homologs in sequenced genomes reveals that many species of bacteria, archaea, and eukaryotes have putative AiiA homologs, most annotated as zinc-dependent hydrolases or β-lactamase-family enzymes. The broad distribution of these proteins suggests that
lactonases activity may be useful in environments ranging from acid mine drainage to agricultural fields.

Among the archaea whose genomes encode an AiiA homolog is *Sulfolobus solfataricus* strain P2, which lives in high-temperature, sulfidic, low-pH environments such as acid mine runoff or volcanic hot springs (49, 53). *S. solfataricus* has been used industrially to leach metals from sulfidic minerals (5, 10, 35, 48), and it has been shown that such leaching is more efficient when the cells adhere to the minerals (21). Although the *S. solfataricus* genome does not seem to encode any HSL synthase or receptor (46), and no report of quorum sensing in *S. solfataricus* exists, the AiiA homolog in *S. solfataricus* may enable this species to compete with HSL-producing species for space on the minerals. In fact, it was recently shown that *Acidithiobacillus ferrooxidans*, which inhabits similar environments, makes HSL signals and attaches to surfaces in response to them (22). In this work, *S. solfataricus* is shown to degrade HSL signals, and SSO1537 is shown to degrade HSLs, most likely as a homoserine lactonase.

**E.2 Methods**

**E.2.1 Bacterial strains and growth conditions.** *Escherichia coli* strain DH10B grown in Luria-Bertani (LB) medium at 37°C was used for manipulation of DNA. The Rosetta strain of *E. coli* (Novagen, Madison, WI) grown either on solid LB or in liquid M9 at 28°C was used for protein expression. When necessary, medium was supplemented with chloramphenicol (25 µg ml⁻¹) or ampicillin (100 µg ml⁻¹). *Sulfolobus solfataricus* strain P2 was purchased from the Deutsche Sammlung von Mikroorganismen und Zellkulturen ([http://www.dsmz.de](http://www.dsmz.de), Braunschweig, Germany) and was grown in Medium 182 at 75°C. *Agrobacterium (A.) tumefaciens* strains NTL1, NTL4(pZLR4) (8), and NTI (pTiC58ΔaccR) (3) were grown on nutrient agar (Difco, Lawrence, KS) or in minimal AB medium (9), supplemented as necessary with gentamicin (25 µg ml⁻¹).

**E.2.2 Identification of an AiiA homolog in *S. solfataricus*.** The amino acid sequence of AiiA protein from *Bacillus cereus* was used as the query term in a blastp search against the non-redundant database at NCBI (2). The predicted protein ORF SSO1537 was aligned with AiiA and other AiiA homologs using ClustalW and a phylogenetic tree was generated using Paup 4.0b (Sinauer Associates, Inc., Sunderland, MA).
E.2.3 Breakdown of homoserine lactones by S. solfataricus. S. solfataricus was grown to an OD$_{730\ \text{nm}}$ ~0.2, then diluted to an OD$_{730\ \text{nm}}$ of 0.04 in Medium 187 with 0 to 10 µM N-oxohexanoyl homoserine lactone (Sigma-Aldrich, St. Louis, MO, catalog # K3007). Replicate tubes without S. solfataricus were also prepared. All tubes were incubated at 75°C for 3 days. After 3 days, AB plates with A. tumefaciens strain NTL4 (pZLR4) and X-gal (0.4%w/v) in an overlay (0.8% agar) were prepared. The S. solfataricus cultures were centrifuged and the supernatants spotted on the overlay plates, which were incubated at 28°C for 3 days. Presence or absence of HSL in the culture supernatants was scored by presence or absence of blue spots on the overlay plates.

E.2.4 Construction of expression plasmid for SSO1537. The primers SSO1537-F1 (GAA ACT TGT TTA TCA TAT GAG TGT AAG T) and SSO1537-R1 (TTA TGA ACT CGA GTT CTT TTG) with the restriction sites NdeI and XhoI (underlined) were used to amplify the open reading frame SSO1537 from S. solfataricus genomic DNA. The resulting PCR product was subcloned into plasmid pTOPO2.1 (Invitrogen, Carlsbad, CA, Catalog # K-4500) to produce plasmid pTA-SSO. The plasmids pTA-SSO and pET16b were then digested with the enzymes NdeI and XhoI. The products of this reaction were separated by agarose gel electrophoresis and the fragments corresponding to linearized pET16b and SSO1537 were excised from the gel, purified with the QIAquick Gel Extraction Kit (Qiagen, Hilden, Germany, Catalog # 28704) and then ligated together to produce plasmid pET16-SSO.

E.2.5 Lactonase activity of SSO1537. After several trials, it was determined that overexpressed SSO1537 was most soluble when the Rosetta cells were grown without induction in M9 medium at 28°C. Replicate culture tubes were prepared with 0, 5nM, 50nM, and 250 nM HSL. The HSL solution in ethyl acetate was dried under N$_2$, then M9 medium and the appropriate antibiotics were added. Last, Rosetta cells with either pET16b or pET16-SSO were added and all cultures were incubated for 24 h at 28°C. 1-ml aliquots were removed from each culture before incubation and after 24 h growth of E. coli. The cells were removed from the aliquots by centrifugation, the supernatant was filtered through 0.2-µm filters, and 0.5 ml of the filtrate was added to cultures of A. tumefaciens at an OD$_{600\ \text{nm}}$ of ~0.05. These cultures were grown for 24 h at 28°C and
aliquots were removed every 4 h and the β-galactosidase activity in the *A. tumefaciens* cells was determined using the Miller assay (36).

**E.3 Results**

**E.3.1 Homologs of AiiA.** A blastp search (2) of the NCBI database shows that the genomes with apparent orthologs to AiiA include fungi, proteobacteria, Gram-positive bacteria, green sulfur bacteria, *Cytophagales* and four species of archaea. These sequences are homologous to, but group separately from, true β-lactamases such as *bla* from plasmid pBR322 (Figure E.2). However, both β-lactamases and HSL-lactonases belong to the zinc-dependent hydrolase superfamily, and have in common a histidine-containing motif (figure E.3).

**E.3.2 Ability of *S. solfataricus* to degrade HSLs.** To confirm that *S. solfataricus* strain P2 is able to degrade HSLs and that the degradation is not due to the culture conditions, *S. solfataricus* was grown with or without *N*-oxohexanoyl-HSL, and parallel cell-free solutions were made with the same concentrations of added *N*-oxohexanoyl-HSL. After growth for 3 days, cells were removed by centrifugation and filtration and the filtrate was spotted on overlay plates with the *A. tumefaciens* indicator strain NTL4(pZLR). The indicator strain has the lacZ gene under the control of traR, the HSL-dependent regulator in *A. tumefaciens*, and will cleave X-gal (and produce a blue color) in the presence of nanomolar quantities of HSL. Blue spots appeared where cell-free solutions containing HSL had been spotted, and where the cultures that had been amended with >250 nM HSL had been spotted. Cell-free solutions without added HSL, cultures without added HSL, and cultures that had started with <250 nM HSL added did not produce a blue color (Table E.1). These results demonstrate that although HSL was still present after 3 days in all cell-free solutions, the cells had degraded the HSL to concentrations below the concentration threshold required to induce of lacZ transcription after the same amount of time.

The supernatants from *S. solfataricus* cultures that had not been amended with any exogenous *N*-oxohexanoyl HSL also did not induce β-galactosidase activity in the detector strain of *A. tumefaciens* (see Table E.1), which shows that *S. solfataricus* does not itself synthesize detectable HSL molecules.
E.3.3 Expression of \textit{SSO1537} in \textit{E. coli}. The gene encoding a putative zinc-dependent hydrolase homologous to AiiA, \textit{SSO1537}, was expressed in \textit{E. coli}. To demonstrate that expression of this gene confers upon this \textit{E. coli} strain the ability to degrade HSLs, \textit{N-oxohexanoyl-HSL} was added to \textit{E. coli} expressing \textit{SSO1537} from pET16b, or to \textit{E. coli} with the corresponding empty vector. Breakdown of the HSLs was monitored by removing 1 ml of the culture before induction and after 24 h of expression, removing the cells by centrifugation and filtration, and adding the supernatant to a liquid culture of \textit{A. tumefaciens} strain NTL4(pZLR4). The amount of LacZ activity in these cells was measured by \( \beta \)-galactosidase assays. After 24 hours of growth of \textit{E. coli} in minimal medium, with or without additional HSLs, the measurable induced \( \beta \)-galactosidase activity was less than half of the \( \beta \)-galactosidase activity induced by the same cultures after 1 h of incubation (Figure E.4).

\textbf{E.4 Discussion}

These experiments have shown that the organism \textit{S. solfataricus} can break down exogenous HSL molecules and that the enzyme encoded by \textit{SSO1537} catalyzes HSL degradation. Although \textit{S. solfataricus} does not produce HSL signals that are detectable by the \textit{A. tumefaciens} strain used here (see Table E.1), it rapidly degrades those added to the culture medium. This is the first reported instance of an archaeon degrading bacterial signaling molecules. In addition, \textit{SSO1537} is the first-characterized archaeal HSL-degrading enzyme as well as the first thermostable representative of the AiiA family.

Expression of \textit{SSO1537} in \textit{E. coli} is sufficient to enable \textit{E. coli} to degrade HSLs. When the mass of the product(s) has been determined, it will be possible to state whether \textit{SSO1537} is a lactone hydrolase, similar to AiiA from \textit{Bacillus} spp., or whether it is an acylase, similar to QuiP from \textit{Pseudomonas aeruginosa}. Because of its sequence similarity to AiiA, especially the conservation of the zinc-binding residues (Figure E.3), it is more likely to be a lactonase. The active site of this protein appears to be very similar to that of the AiiA protein that was recently crystallized from \textit{Bacillus thuringiensis} subsp. \textit{kurstaki} HD263 (Kim \textit{et al.}, 2005). The zinc-binding residues in the active site of AiiA are absolutely conserved in all four of the archaeal homologs (see Figure E.3), and
the *S. solfataricus* AiiA amino acid sequence is 29% identical and 47% similar to the *B. thuringiensis* sequence overall.

Similar putative zinc-dependent hydrolases appear in the genomes of a variety of primarily bacteria, but some eukaryotic and archaeal genomes. Among these bacteria are four species of green sulfur bacteria (GSB), including *Chl. chlorochromatii*, the GSB epibiont of the consortium “Chlorochromatium aggregatum” (50). The role of this protein in such a consortium could be interesting. Approximately 20 GSB surround a central β-proteobacterium, attached by structures which cross the membranes of both species (50). Neither those genes required for synthesis of HSLs nor those encoding HSL receptors have been identified in the genomes of any GSB. In fact, due to the small number of predicted transcriptional regulators in GSB genomes, it has been assumed that they regulate only a few processes at the level of transcription (20). However, it is possible that in the case of *Chl. chlorochromatii*, the ability to degrade exogenous HSL signals might be beneficial. Perhaps the GSB are preventing the central β-proteobacterium from escaping. “Chlorochromatium aggregatum” has been shown to be phototactic as well as chemotactic towards organic carbon compounds and sulfide (23); perhaps breakdown of the HSL molecules in the GSB component of the consortium starts a signaling cascade that transmits information to the central rod.

The occurrence of AiiA homologs in several species of archaea (see figure E.2) and the demonstrated ability of *S. solfataricus* to degrade HSLs extends “quorum-quenching” activity (16) into a third domain of life. In any environment limited in some nutrients, such as high-temperature sulfidic springs, acid mine drainage, or solfatara fields, the organisms living there must compete for resources, especially if the species there have similar nutritional requirements. The coexistence of species communicating via HSL signaling and species that degrade those signals may simply mean that an equilibrium exists between the two populations, which allows both to establish stable populations. Breakdown of the signals used by another species as a cue to colonize could prevent those species from invading locations already occupied by *S. solfataricus*, while (because AiiA is not predicted to be an extracellular protein) leaving the unoccupied areas free for those bacteria to attach. It is also possible, though less likely, that the degradation of HSL signaling molecules by *S. solfataricus* reduces the concentration of
the signals in the environment to levels sufficiently low that the species using those signals cannot detect them, and thus cannot coordinately regulate their activity.

Alternatively, S. solfataricus may, like the alga Ulva intestinalis, only colonize surfaces from which HSLs are diffusing (29, 42, 47). If the bacterial HSL-producing species grow in biofilms where nutrients are plentiful, or if in the course of their metabolism they release compounds that can be used as carbon or electron sources, the HSL signals may be a chemoattractant to S. solfataricus. Many species of Sulfolobus have flagella (25), at least one species has been shown to swim in response to environmental cues (31), and the S. solfataricus strain P2 genome encodes at least one methyl-accepting chemotaxis protein and accessory flagellar proteins (46). If the HSL-synthesizing bacteria have nutrient requirements similar to those of S. solfataricus, it is possible that S. solfataricus could swim along an HSL concentration gradient to find a surface suitable for attachment and growth.

It is also possible that S. solfataricus does not compete with HSL-synthesizing bacteria, but cooperates with them. In some acid mine drainage systems, pure cultures of organisms isolated from those environments cannot break down minerals, but co-cultures of two or more break them down efficiently, freeing nutrients from insoluble mineral grains (e.g. 1, 6, 28, 40). As suggested above, HSL molecules could be chemoattractants for S. solfataricus, which may subsequently degrade the signals for use as a carbon source.

The characterization of an HSL-degrading protein produced by the hyperthermophile S. solfataricus opens up a series of interesting questions regarding inter-domain signaling, cooperation, and competition. With the existence of a genetic system for S. solfataricus (11), it should now be possible to study the role of AiiA in this organism both in pure culture and in co-cultures with HSL-producing bacteria.

Acknowledgements
The E. coli Rosetta cells were the gift of Dr. J. Gregory Ferry, The Pennsylvania State University. The Agrobacterium tumefaciens strains were the gift of Dr. Steven K. Farrand at the University of Illinois at Urbana-Champaigne. Funding for this project was provided
References


sensing regulator controls expression of both the Type IV secretion system and the flagellar apparatus of *Brucella melitensis*. Cell. Microbiol. 7: 1151-1161.


Figure Legends.
Figure E.1. Hydrolysis of acyl-homoserine lactone to acyl-homoserine. The product of this reaction does not bind to the HSL receptors in the cells and is inactive as a signaling molecule.

Figure E.2. Phylogenetic tree of AiiA homologs found in different species. Bold lines indicate proteins whose activity has been characterized biochemically. The outgroup is the β-lactamase from pBR322.

Figure E.3. Alignment of AiiA homologs from archaea with the well-characterized AiiA from *Bacillus* spp. The zinc-binding motif is highlighted by a box. Residues marked with red stars bind Zn in the crystal structure of AiiA from *Bacillus thuringiensis*.

Figure E.4. Maximum β-galactosidase activity measured in *A. tumefaciens*. β-galactosidase activity in *A. tumefaciens* is activated by addition of conditioned medium from the *E. coli* cultures with or without SSO1537. The additions are conditioned medium from *E. coli* cultures before induction of SSO1537 expression (A) and conditioned medium from *E. coli* cultures after 24 h of expression at 28°C (B).
**Table E.1** Development of blue color in *Agrobacterium tumefaciens* overlay plates. Blue color indicates cleavage of X-gal by LacZ, which is under the control of the quorum-sensing system in this strain of *A. tumefaciens*. (+) indicates appearance of blue color; (-) indicates that no color developed. SSO, *S. solfataricus* cultures grown with the initial HSL concentration indicated in the left-most column. CF, cell-free controls inoculated with the same HSL concentrations and incubated under identical conditions as the *S. solfataricus* cultures.

<table>
<thead>
<tr>
<th>[AHL]</th>
<th>SSO.1</th>
<th>SSO.2</th>
<th>SSO.3</th>
<th>CF.1</th>
<th>CF.2</th>
<th>CF.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 µM</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>10 µM</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>5 µM</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>5 µM</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>1 µM</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>1 µM</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>0.5 µM</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>0.5 µM</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>0.1 µM</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>0.1 µM</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>50 nM</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>50 nM</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>10 nM</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>10 nM</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>5 nM</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>5 nM</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Figure E.1
Figure E.2

- Bacillus cereus AiiA
- Bacillus thuringiensis AiiA
- Bacillus subtilis AiiA
- Bacillus sp. BS4 AiiA
- Klebsiella pneumoniae AhIK
- Agrobacterium tumefaciens AttM
- Arthrobacter sp. AhlD
- Sulfolobus sulfataricus AiiA
- Aspergillus clavatus AiiA(h)
- Aspergillus oryzae AiiA(h)
- Halobacterium sp. NRC1 VNG1340C
- Chlorochromatium aggregatum Cag_1138
- Prosthecochloris aestuarii Paes_0740
- Salinibacter ruber Sru_1205
- beta-lactamase, pBR322
Figure E.4

A. 

B. 

maximum $\beta$-galactosidase activity

pet16  pSSO

o AHL  5nM AHL  50nM AHL

maximum $\beta$-galactosidase activity

pet16  pSSO

o AHL  5nM AHL  50nM AHL
APPENDIX F

COMPLETE CURRICULUM VITAE
Julia A. Maresca  
Department of Biochemistry and Molecular Biology  
The Pennsylvania State University  
University Park, PA 16802  
e-mail: jam636@psu.edu  
jam636@psu.edu  

Education  
• 2007: Ph.D., Department of Biochemistry and Molecular Biology, The Pennsylvania State University, State College, PA.  
• 2002: Microbial Diversity Summer Course, Marine Biological Laboratory, Woods Hole, MA.  
• 1999 B.A., Biology, University of Chicago, Chicago, IL.  

Research Experience  
• 2001-2007: Graduate Research Assistant, Department of Biochemistry and Molecular Biology, The Pennsylvania State University. Advisor: Dr. Donald A. Bryant.  
• 2000-2001: IRTA Post-Baccalaureate Fellow, Lab of Viral Diseases, National Institute on Allergy and Infectious Disease. Advisor: Dr. Alison McBride.  
• 1998-1999: Undergraduate researcher, Department of Organismal Biology and Anatomy, University of Chicago. Advisor: Dr. Martin Feder.  

Publications  


Fellowships and Awards

• 2007. Graduate School Alumni Association Dissertation Award, The Pennsylvania State University, for “The genetic basis for pigment variation in green sulfur bacteria.”
• 2005. International Travel Award, Biogeochemical Research Initiative for Education (NSF-IGERT)
• 2004-2006. NASA Space Grant Fellowship, Pennsylvania Space Grant Consortium.
• 2004. Summer Research Fellowship, Center for Environmental Chemistry and Geochemistry, The Pennsylvania State University, for “Nutrient cycling in a stratified lake.”
• 2003. Research Award, Biogeochemical Research Initiative for Education (NSF-IGERT) for “Breakdown of bacterial homoserine lactone signals by archaea.”
• 2002-2004. NASA Space Grant Fellowship, awarded by the Pennsylvania Space Grant Consortium.

Abstracts submitted to conferences


2. J.A. Maresca, T. Li, D.A. Bryant. Identification of a gene cluster responsible for the brown phenotype in green sulfur bacteria. Eastern Regional Photosynthesis Conference, Woods
Hole, MA., April 2006. (Oral presentation)

3. **J.A. Maresca**, D.A. Bryant. Interspecies signal abduction: An archaeal protein that degrades homoserine lactones. Environmental Chemistry Student Symposium, March 2006. (Poster; Winner, Best Poster Award)

4. C.M. Cress, **J.A. Maresca**, and D.A. Bryant. Characterization of the transcriptional regulator ModE in *Chlorobium tepidum*. Environmental Chemistry Student Symposium, March 2006. (Oral presentation; Winner, Best Undergraduate Student Presentation)


7. **J.A. Maresca**, T. Li, D.A. Bryant. Using comparative genomics to identify unique genes in bacteriochlorophyll biosynthesis. Allegheny Branch of the American Society for Microbiology annual meeting, October 2005. (Oral presentation; Second Prize, Graduate Student Oral Presentations)


**Teaching Experience**

- Supervising honors student research (BMB 496). Oxygen tolerance of a suite of carotenoid mu-

- Supervising summer undergraduate research. Molybdenum dependence of nitrogenase expression in *Chl. tepidum* and characterization of *Chl. phaeobacteroides* strains in Fayetteville Green Lake, NY. Lindsay Glace and Jennifer Vrentas, Summer 2004.


- Lab instructor, Department of Biochemistry and Molecular Biology, Pennsylvania State University


**Outreach**

- Organizer, brown-bag lunches for women in BMB Department. Fall 2005.
- Member, Organizing Committee, Environmental Chemistry Student Symposium. 2004.

Julia A. Maresca

Education
• 2007: Ph.D., Department of Biochemistry and Molecular Biology, The Pennsylvania State University, State College, PA.
• 1999 B.A., Biology, University of Chicago, Chicago, IL.

Selected Publications

---

1 For a complete CV, please see Appendix F.